

A sound approach to GM debate

If genetic modification is to yield benefits in socially acceptable ways, governments need to ensure that there is broad but well-focused consultation. A New Zealand commission provides an excellent example.

Few areas of science and technology touch as directly on the lives of citizens as genetic research. So it is not surprising that scientists working at its frontiers find their enthusiasm for the technological and social possibilities challenged by a tide of public opinion (real or apparent) and ambivalent political responses. Whatever researchers may believe about the benefits, the future of genetically modified (GM) crops and foods depends on the prosperity of companies wishing to invest in their development and on the willingness of farmers, retailers and consumers to buy them. Those market forces in turn depend critically on regulation and public attitudes.

Thus it has been encouraging to witness the constructive and sensitive approach adopted by the New Zealand government in establishing a Royal Commission on Genetic Modification as applied to research, medicine and agriculture. Its scope and processes have been unique, as justifiably claimed by New Zealand's prime minister, Helen Clark, on releasing its valuably comprehensive report last week (see page 573).

According to the report, "genetic modification" was to be addressed in all its ramifications, not overtly for the global interest, but for New Zealand's better management of its own immediate challenges. Yet the report did not consider the issues in South Pacific isolation. The inquiry cast its net for witnesses worldwide, and merits international readership (see <http://www.gmcommission.govt.nz>). In the end, a campaign to make New Zealand a genetic-engineering-free zone has, with transparent justice, been dealt a heavy blow from which it will be difficult for it to recover, although New Zealand's Green Party vows to "fight on".

Ambition rewarded

The commission's remit was an ambitiously broad one for four people to tackle in little over a year. They were charged with covering the whole gamut of scientific, economic, environmental, ethical, indigenous, intellectual-property, legislative and regulatory angles of safety and risks of genetic modification.

New Zealand is a compact society, and its economy is heavily dependent on the competitive export of agricultural produce. In an economy that has been slipping steadily in the world rankings, there has been a strong push for genetic research and application to become the saviour that delivers enhanced productivity.

On the other side has been a politically vociferous Green movement, with seven members of parliament who theoretically hold the balance of power over the Labour-Alliance coalition. The Greens successfully pressed for the establishment of the Royal Commission. This is a form of quasi-judicial inquiry in countries that inherited British law which ensures its independence of government and institutions and confers powers not available to other forms of inquiry, in particular the cross-examination of witnesses.

Taking account of the sensitivities of Maori, the indigenous community comprising about 12% of the population of 3.9 million and with special rights under the nineteenth-century Treaty of Waitangi, was an overriding condition. That need resonates with many other nations in resolving questions of the ownership of native biota.

The commission chairman and former chief justice, Sir Thomas Eichelbaum, guided the broad involvement by general public and experts alike, beginning with 'scoping' meetings to define the issues more clearly. Eichelbaum is proud of having incorporated social and indigenous values into the management of genetic science: "Few minds may have been changed in the process but everyone emerged better informed and more willing to listen to each other."

Wide consultation

The commissioners consulted widely—and always openly—through 15 public meetings and three forums across the country, including special emphasis on Maori, with whom they held 11 "hui" (community gatherings). Evidence was taken from 107 "interested persons", who first had to qualify under the Commissions of Inquiry Act. Some of them presented their cases with legal backing, and witnesses were cross-examined by commissioners, their legal counsel and opposing parties. But this was not conducted in the adversarial manner that scientists called as expert witnesses in court rightly detest.

For a small country, the commission attracted a huge number of submissions (10,861). These were overwhelmingly against genetic modification (92%, and 65% strongly so). But an extensive public survey found a different balance of opinion. Prompted to nominate important issues, only 2% mentioned GM concerns. The report is replete with detailed analyses of submissions that ultimately reinforced the commissioners' decision to reject the Greens' key policy.

Jean Fleming, a reproductive biologist at the University of Otago, found that her fellow commissioners were influenced in the end by the greater weight of scientific evidence over assertions about risk. She perceived that there were few entrenched 'anti-scientists' among participants, as all sides called for more research on areas of doubt. However, a lack of trust of scientists was evident, with many blaming the influence of the profit motive forced on researchers in New Zealand since a massive restructuring of public science ten years ago. Fleming was saddened by the fact that the submissions of several scientists, on whose evidence the Greens were basing their case of unacceptable risk, fell apart under cross-examination.

New Zealand's consultation stands in markedly favourable contrast to its neighbour's approach. In Australia, communications to and from the public are less clear, with five government departments involved and none being advised by consultative processes remotely similar to the New Zealand inquiry.

Having established a model of community consultation and scientific rigour that other nations may consider emulating, the New Zealand government cannot rest on its laurels. Some of the commission's recommendations require further public resources. It is all too easy to request more funds for research, but the commission is surely right to highlight the need for publicly funded exploration of the environmental impact of GM crops as well as research into organic and other sustainable agricultural systems. But the report's recommendations are much more wide-ranging and, in places, contentious. To consolidate the commission's good work, the New Zealand government will need to legislate with determination. ■





Canadian club
New national body
could become voice
of science in Canada
p573



Top scorers
Soccer-playing
robots approach
the AI goal
p574



Hot debate
Congress urges Bush
to set limits on
carbon emissions
p575



Prized position
Italian Nobellist
appointed senator
for life
p576

Mad-cow outbreak spurs German drive to combat prion diseases

Alison Abbott, Munich

German biologists and veterinarians are preparing to launch a coordinated investigation into prion-based diseases, which include bovine spongiform encephalopathy (BSE), or mad cow disease.

At a government-sponsored meeting in Munich on 20 July, researchers met to determine questions that should be asked and the necessary tools for studying these diseases, of which Germany has little direct experience.

Because Germany was thought to be free of BSE, German research on the diseases, known collectively as transmissible spongiform encephalopathies (TSEs), figured little during the 1990s.

But since the first cases were discovered there last year (see *Nature* **408**, 506; 2000), sources of research funding have been springing up. Sponsors include consumer-protection and research agencies at both state and federal level. Earlier this year, for example, three states earmarked special research funds for prion-based diseases. In May the federal government allocated DM27 million (US\$12 million) to such research.

A significant part of the federal funding will support a "research platform" to coordinate the mushrooming research and provide infected material for study.

"So much has been happening so fast that it is hard to follow — and coordinate with —



Research model: Martin Groschup wants to harmonize German and French efforts on BSE.

initiatives in other parts of Germany," says Hans Kretzschmar, a neuropathologist at the Ludwig Maximilian University of Munich.

Aside from sponsoring meetings such as last month's, the government will create a web-page repository for results and resources, including material from a federally funded pathogenesis study due to begin at the end of

the year. The study will be based at the new Institute for Emerging and Unconventional Infectious Agents on the remote Baltic island of Riems, north of Greifswald, where 50 cattle will be deliberately infected with BSE. As TSE research increases, "German scientists will need a more direct supply of infected tissue", explains Martin Groschup, a veterinary virologist at the Federal Research Centre for Virus Diseases of Animals in Tübingen, who heads the pathogenesis programme.

Regular samples of biological fluids, including blood, cerebrospinal fluid and urine, will be taken from the cattle, and a small number will be slaughtered every few months for complete dissection. The tissues will be distributed to TSE researchers.

Groschup hopes to harmonize the programme with a similar exercise being proposed in France. If the French programme materializes, he says, "we want the inoculum and procedures to be standardized so that results from research on the materials are truly comparable."

Kretzschmar and Groschup are spearheading the TSE research programme in Bavaria, which, with DM20 million of funding over two years, is Germany's largest such state programme. Bavaria is also building a DM38-million TSE research centre in Munich for work on infected animals, particularly mice. Two other states, Baden-

Drugs firms inflate research costs, watchdog says

Jonathan Knight

A watchdog group has challenged the US pharmaceutical industry's claim that drugs cost an average of \$500 million to develop. The actual cost, according to a report from Washington-based Public Citizen, is closer to \$100 million. The group charges that industry lobbyists have misled Congress and the public to justify high prices.

According to the report, financial risks are lower because companies cherry-pick research from publicly funded laboratories. An internal document obtained by Public

Citizen from the National Institutes of Health estimates that over half the research that led to development of the top five new drugs in 1995 came from the public sector.

But this ignores the gap between lab bench and pharmacy, says Iain Cockburn, an economist at Boston University. The report's authors "are too ready to discount the risk of trying to turn promising lab research into a marketable drug", he says.

Public Citizen claims that the often-cited cost of \$500 million ignores tax breaks on research and development. Bob Young, who

directed the project, says it is also inflated by a theoretical value for the 'opportunity cost' of capital, or what it would have earned if invested elsewhere. But Joseph DiMasi, the economist at Tufts University in Boston on whose 1991 study the higher figure is based, says: "Opportunity cost is a real factor."

Public Citizen has called on Congress to institute price caps for prescription drugs supplied by Medicare, the federal insurance programme for retired people.

► <http://www.citizen.org/congress/drugs/r&dscarecard.html>

Württemberg and Niedersachsen, have smaller programmes of their own.

The upsurge in funding has raised eyebrows among health researchers specializing in more common diseases. The human TSE — variant Creutzfeldt–Jakob disease (vCJD) — has so far claimed only around 100 victims, none of them in Germany.

“CJD is not a big health hazard, but it is worth funding generously,” says Kretzschmar. This is not just because it raises hopes of curing a rare disease, he argues, but also because it feeds into research on neurodegenerative diseases generally, which usually involve amyloid proteins. “The only difference between these diseases and diseases such as Alzheimer’s is that TSEs are transmissible,” he says.

One of the Bavarian projects involves screening chemical libraries for their ability to prevent prion proteins from aggregating, using a new screen that tracks the diffusion time of a single fluorescently tagged prion molecule. This work will be done by academics because, says Kretzschmar, “the pharmaceutical industry is not interested in finding cures for vCJD because it is so rare”.

Another project is to create genetically modified cattle that lack the gene for the prion protein PrP, and which therefore cannot contract BSE. ■



Hans Kretzschmar
says prion research
is worthwhile.

Medical journals seek means to free authors from industry

Paul Smaglik, Washington

Several leading medical journals are planning a new publishing policy designed to empower academic authors who collaborate with industry.

Aimed at authors who work with drug companies on clinical trials, the policy should help academics retain full control of the content and timing of research results produced in industrial collaborations.

Critics contend that drug companies can exert excessive influence over the publication of work that they have paid for, potentially suppressing negative results, for example.

The journals — including the *Journal of the American Medical Association (JAMA)*, the *New England Journal of Medicine* and *The Lancet* — will announce the joint policy in mid-September. They are expected to reject reports from trials sponsored by drug companies unless the authors have been granted explicit control over the data and the decision to publish.

In an interview with *Nature*, Jeffrey Drazen, editor-in-chief of the *New England Journal of Medicine*, was tight-lipped about the details of the policy, whose existence was reported in the *Washington Post* on 5 August. But he was less coy about the circumstances that prompted the editors to develop it.

“Academic investigators have had less and less opportunities to work with pharmaceutical sponsors with respect to study design,

data analysis and interpretation, and manuscript drafting,” he says. “They’re given a ‘take it or leave it’ stance.”

The new policy is intended to help academic investigators negotiate more favourable terms with companies, Drazen says. As companies seek the endorsement that publication of a positive trial in a journal implies, they may be willing to give more freedom to their academic collaborator.

George Lundberg, editor-in-chief of online medical journal Medscape, says that the policy’s overall goals are “praiseworthy”. But he is unsure how effective it will be in helping journals establish more editorial independence. “The devil will be in the details,” he says.

For example, Lundberg wonders how far any policy where authors have to pledge their independence can be extended, and whether authors will be expected to vouch for their full independence from their universities and granting agencies as well.

News of the policy attracted a mixed reception. A spokesperson for the Pharmaceutical Research and Manufacturers of America questions whether the policy is necessary. But the New York-based Citizens for Responsible Care & Research, a consumer group that has been critical of medical research conduct, says that it supports the policy’s ends, but doubts that it will contain the means to achieve them. ■

Golf course threatens to leave hole in fossil records

Rex Dalton, San Diego

A planned golf course outside Denver, Colorado, could swallow up a valuable dinosaur track site that includes the world’s only known tracks of a species from 65 million years ago, palaeontologists say.

According to the plans, some of the tracks are to be covered with dirt, and some cut out of their stone and removed for display elsewhere. Others would be preserved *in situ* amid the fairways and greens of the proposed municipal course in the small town of Golden in the Rocky Mountains.

But the plans have met with fierce opposition from geologists and palaeontologists. William Caneer, a retired geologist from the Colorado School of Mines, says: “Palaeontologists spend half their lives digging specimens up; now the city wants to cover tracks with dirt. It makes smoke come out of my ears.”



Rough justice? Many of the fossils at Golden could be covered over by fairways and greens.

Colorado’s state archaeologist, Susan Collins, has requested a survey of the palaeontology specimens at the site known as the Parfet clay pits, a former clay mine on the eastern edge of the Rocky Mountains.

Collins says her agency will determine what can be done with the track site after the

survey is complete. But she has not ruled out accepting the plan for the golf course.

Once a waste dump for coal-plant ash, and broken bottles from the nearby Coors Brewery, the privately owned clay pits — recently donated to the city of Golden — were for years the domain of amateur rock hunters, who cut out tracks with high-powered rock saws.

Around 1985, Martin Lockley, a geologist at the University of Colorado at Denver, and colleagues found in the pits the first ever identified tracks of a ceratopsian, along with the only known US track of the crocodile-like champsosaur. Golden has contracted Lockley to assist in preserving the tracks during construction of the golf course.

Lockley, who curates the world’s largest dinosaur-track collection at his university, says the course could incorporate the history of the site by including appropriate displays of the tracks. ■

Canada plans to give unified voice to science

David Dickson

Canada is planning to establish a single national science organization, designed to generate and coordinate scientific advice to the federal government and to provide a voice for Canadian science in national and international debates.

Draft proposals for such a body — to be known as the Canadian Academies, and based on models such as the US National Academies complex — are now being published on the . by a working party set up by the federal government last year.

The working party describes the establishment, which would be run as a non-profit charitable organization with its own president and full-time staff, as an “imperative” for the country, given the growing social and political importance of scientific issues.

Its members would include the country's three main existing organizations for science, engineering and health: the Canadian Academy of the Sciences and Humanities (otherwise known as the Royal Society of Canada), the Canadian Academy of Engineering, and the Canadian Academy of Health Sciences, which is to be established later this year.

One enthusiastic proponent is the chair of the working party, Gilbert Normand, Secretary of State for Science Research and Development. “The Canadian government has a large number of separate advisory committees, but it does not have an independent, national organization that has the confidence both of the Canadian people and of the international scientific community,” he says.

According to Normand, one of the main tasks of the new body will be to provide a source of “credible, independent expert assessments on the sciences underlying important issues and matters of public interest”. To carry out these assessments, the body will use either its own money or funding from the government or other sources.

The working group, which includes Michel Chrétien, director of the Regional Protein Chemistry Centre at the Ottawa Health Research Institute and brother of Canadian Prime Minister Jean Chrétien, has planned in detail how the new body might operate.

One suggestion is that, as well as members selected by each of the participating organizations, its board of governors would include six members appointed from the public.

Normand says that, providing that there is general public support for the plans — which have met no opposition so far — he intends to propose the creation of the Canadian Academies to the cabinet “sometime in the autumn”, and that “perhaps there will be some money allocated in the next budget”.

Running costs are estimated to be Can\$3 million a year. Normand says that, although the government might decide to provide this



All for one: Canada's parliament stands to benefit from scientific advice given by the proposed body.

money on an annual basis, his preferred option would be to set up the organization with an initial capital allocation from the government of Can\$30 million, which would allow it to be stable for ten years of operation and would help to nurture its independence.

Another keen supporter of the project is Tom Brzustowski, president of the Natural Sciences and Engineering Research Council

of Canada, who says that the idea first came to him after the World Conference on Science in Budapest in June–July 1999.

Shortly after the conference, Brzustowski wrote in the research council's newsletter that he felt that Canada lacked the institutional capacity to deal with the “big issues” involving science and society.

► <http://www.nrc.ca/indcan/nso>

Commission plots transgenic future

Peter Pockley, Sydney

An extensive study by a royal commission has opened the door for New Zealand to cautiously embrace genetically modified (GM) agriculture for the first time.

The findings of the Royal Commission on Genetic Modification in New Zealand were welcomed by many scientists. But they angered the country's Green Party, whose considerable political influence led to the commission being established.

Peter Gluckman, dean of medicine at the University of Auckland, says the report is “very sensible in that it rejected outright the concept of a genetic-engineering-free New Zealand as incompatible with the modern world and the nation's future”.

The commission's 1,200-page report, released on 30 July, says that transgenic agriculture should be introduced to New Zealand “selectively with appropriate care”. It recommends loosening existing controls on field trials of GM crops, and creating new mechanisms for controlling their commercial release. No GM crops have yet been released for commercial sale in New Zealand.

But the commission says that genetic modification should be banned in certain circumstances where its introduction

might threaten growers' interests.

It also suggests that decisions on the first commercial release of GM crops should be shifted from the Environmental Risk Management Authority to the environment minister. A parliamentary commissioner on biotechnology would be given powers to investigate issues, independently of the government, to produce accessible reports for the public and to advise parliament on genetic-engineering policies.

The royal commission was established by the Labour–Alliance government 15 months ago (see *Nature* 404, 914; 2000). It is made up of four commissioners — a retired judge, a biomedical researcher, a medical practitioner of Maori heritage and an Anglican bishop.

The government is not bound by the commission's recommendations, but the prime minister, Helen Clark, and the environment minister, Marian Hobbs, welcomed the commission as “the most wide-ranging inquiry into genetic modification ever undertaken in any country”. They set a deadline of 31 October for announcing the government's plans to enact its recommendations.

► www.gmcommission.govt.nz

Whistle-blowers wait for overbilling verdict

Rex Dalton, San Francisco

A federal judge will decide next week whether payments should be made to whistle-blowers who expose improper government billings made by public medical schools in the United States.

The San Francisco judge will hold a hearing on 16 August to determine whether four whistle-blowers should share part of \$22.5 million paid by the University of California to settle lawsuits alleging that government health-care programmes were overbilled by hospitals at the university's five medical schools.

The four women filed the lawsuits several years ago under the federal False Claims Act, which permits a whistle-blower to receive 15–25% of the funds the gov-

ernment recovers after exposure of fraud or abuse.

The lawsuits alleged fraudulent billing practices on behalf of University of California faculty, who may have been involved in research, consulting or other activities when they were supposed to be supervising patient care by training physician residents.

The case will be watched closely by researchers at US medical schools. A decision supporting continued close government scrutiny of billing practices will be viewed by some physicians as inhibiting their ability to conduct research. But some non-physician researchers may welcome such scrutiny, which they think serves to level the playing field for all researchers in the medical schools.



The US Department of Justice is asking the judge to deny payment to the whistle-blowers, who are seeking \$4.5 million of the settlement money that the university paid earlier this year to the government.

If the judge decides the whistle-blowers should not receive a portion of the settlement, both sides agree that the decision could prevent any future false-claims lawsuits against public medical schools. Private universities are not affected by the case.

Stephen Meagher, a San Francisco attorney representing two of the whistle-blowers, says the government has turned its back on his clients after they exposed extensive improprieties at the university. "The government wants people to expose abuse," he says. "But they don't want to pay them; it won't work."

Federal attorneys say the government does not have to share the settlement with the whistle-blowers because of a US Supreme Court decision last year limiting False Claims Act lawsuits against public entities, such as state universities.

Government audits of a five-year period to 1998 found that more than \$200 million in excess payments were made to the University of California medical schools, a university attorney said. But the university disputes the audit conclusions, and admitted no wrongdoing in agreeing to pay the \$22.5 million.

Late last year, the government discussed a \$1 million payment to the whistle-blowers, Meagher says, but halted negotiations shortly after the Republican administration of President George W. Bush took office. Some Republicans have voiced scepticism about rewarding whistle-blowers and their lawyers for exposing the misuse of public funds, but they have been unable to repeal the False Claims Act.

The Association of American Medical Colleges, which represents the medical schools, asserts that its members have been unfairly singled out for government audits.

Soccer robots get the ball rolling

Quirin Schiermeier

The fifth world championships for autonomous soccer-playing robots kicked off in Seattle, Washington, last week at an international meeting of artificial intelligence (AI) researchers. A record number of 120 teams from 25 countries will compete in the RoboCup, which is being held in the United States for the first time.

Soccer is the simplest of sports. But the number of human decisions, physical forces and interactions between them — its number of degrees of freedom — is immense. So it is no surprise that AI researchers and software engineers have homed in on the world's most popular sport as a promising arena for their work.

The tournament will provide them with ample opportunity to demonstrate, compare and improve the complex coordination, navigation and decision-making abilities

of different robots. And scientists and spectators alike can get a kick out of it.

The navigational and cognitive skills needed to play soccer are also useful for developing robots to save humans from dangerous situations such as fires or earthquakes, the researchers say. Prototypes of rescue robots and two-legged robots will be on show in Seattle.

Whereas the first generations of robots could only recognize obstacles and get out of their way, today's intelligent machines can perform relatively complex tasks requiring division of labour, make informed choices, and learn from their mistakes.

Whether robots will ever play soccer with any real proficiency remains to be seen. But the organizers of the RoboCup claim that a team of robots will beat the human world champions by the year 2050.



Booting up: the University of Freiburg team (left), which won the German Open in June, is defending its world title at RoboCup 2001 in Seattle — where improving artificial intelligence is the goal.

Senate urges Bush to act on climate change

Quirin Schiermeier

Despite President George W. Bush's rejection of the Kyoto Protocol on climate change, the US administration is under mounting pressure from Congress to constrain the country's greenhouse-gas emissions.

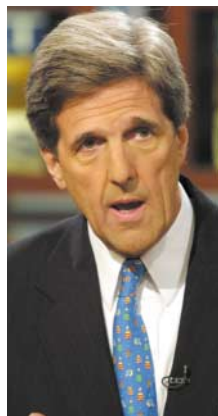
On 1 August, a powerful senate committee called on the Bush administration to secure the participation of the United States in a revised Kyoto Protocol or other future binding agreement on climate change. The call was made unanimously by the Senate Committee on Foreign Relations.

And on 3 August, two influential senators, John McCain (Republican, Arizona) and Joseph Lieberman (Democrat, Connecticut), jointly announced that they would propose legislation later in the year to set mandatory limits on the United States' greenhouse-gas emissions.

The 19–0 vote by the foreign-relations committee surprised some observers, as several of its members, including Charles Hagel (Republican, Nebraska), have strongly criticized the Kyoto Protocol in the past.

The resolution was proposed by John Kerry (Democrat, Massachusetts) as a non-binding amendment to a bill funding the Department of State. Its main goal is to bring the United States back to the negotiation table, Kerry says.

The resolution calls on President Bush to put forward a proposal at October's meeting in Marrakesh, Morocco, of the parties to the Kyoto accord in an attempt to secure the United States' participation in a revised



Beating the gridlock: John Kerry hopes the United States will resume talks on carbon emissions.

protocol or other climate-change agreement.

A 1997 Senate resolution, which said that developing nations need to be included in any emissions treaty, should not cause the United States to abandon "its shared responsibility to help find a solution to the global climate-change dilemma", the resolution adds.

Environmental groups praised the resolution, saying that it reflects growing momentum in the Congress for action on climate change. Elliot Diring of the Pew Center on Global Climate Change, a group that encourages dialogue between industry and the public on the issue, notes that the resolution is the first time the Senate has declared that the United States should be party to a mandatory treaty on climate change.

But the Global Climate Coalition (GCC),

which coordinates business opposition to mandatory controls of greenhouse-gas emissions, claims that the new resolution affirms Bush's opposition to the Kyoto accord. "The amendment effectively made the case yet again that any international approach to climate change must involve commitments from developing nations and must not harm American families, workers and communities," says Frank Maisano, a GCC spokesman.

Almost 180 governments signed up to revised Kyoto rules in Bonn last month (see *Nature* **412**, 365; 2001). But the United States — which emits about 25% of the world's greenhouse gases — rejected the agreement, saying it would unacceptably weaken the country's economy, and that the treaty should not exclude developing countries from binding reduction targets. ■

Enterprising drug company offers cash for chemicals

David Adam, London

Feeling short of funds for that convertible sports car or Caribbean holiday? For those with the right blend of chemistry skills, mountains of cash are now available: all you have to do is synthesize some valuable but awkward molecules on a freelance basis.

Eli Lilly, the Indianapolis-based pharmaceutical company, is offering cash rewards of up to \$100,000 to researchers who can help it to solve bottlenecks in its drug-discovery programmes. The company has posted a list of more than 20 molecules on the Internet and is asking anyone with the know-how — from postgraduate students to retired university professors — to come up with better ways of making them.

For example, the company needs a new way to synthesize a particularly tricky bicyclic ethyl ester. Avoid "highly toxic

reagents and solvents" and stay within "–30 °C to 150 °C" and you could earn \$80,000 — if you can deliver several grams of the ester by 31 October and your method is judged to be the best. Researchers who do not have access to laboratory facilities can also have a go at cracking the conundrums: paper solutions are being accepted for some of them.

Since launching its new web-based venture — called InnoCentive — last month, Lilly says it has had an "extraordinary" response. But the problems clearly leave many would-be contestants scratching their heads: the 900 or so people who have registered for full details have so far submitted just a handful of solutions between them.

The company expects to award its first cash prize early next month, says Alpheus Bingham, vice-president of research and development for Lilly's venture-capital

arm, who is heading the project.

Adam Nelson, an organic chemist who specializes in asymmetric synthesis at the University of Leeds in Britain, says the problems do not look too tricky. "They're not fiendishly difficult," he says. "In fact, they're the kind of problems that an average PhD student will be solving a lot of the time. But clearly there's a methodology gap here." Some lab-scale solutions will prove difficult to scale up to produce the large quantities needed for industry, for example.

Nelson plans to set one of his students to work on one of the problems: a substituted thiophene with a \$70,000 reward. "It's exactly the sort of chemistry that we do," he says. Researchers in other fields should also stay tuned: Bingham says that more problems, this time in biology and informatics, will be posted soon. ■

► <http://www.innocentive.com>

Panel gets to grips with soaring costs of space station

Washington NASA administrator Dan Goldin has asked an external review panel to investigate cost overruns on the International Space Station. The over-spending is threatening to damage the facility's usefulness as a research laboratory (see *Nature* **412**, 465; 2001).

The panel, made up of engineers, business executives and scientists, will be chaired by Thomas Young, former president of the Martin Marietta Corporation, an aerospace and defence company that is now part of

Lockheed Martin. Other members include heart surgeon Michael DeBakey and two Nobel laureates, physicist Robert Richardson and genetics researcher Richard Roberts.

Young also chaired the panel that examined the loss of two NASA Mars missions in 1999, and already heads the agency's standing committee on human spaceflight, which includes the space station. The panel is due to report its findings to Goldin on 1 November.

NASA is currently looking at a number of ways of cutting the cost of the ISS. These include scrapping key pieces of lab equipment due to be installed after 2004 and abandoning work on a space-station rescue vehicle. This latter measure would limit the number of crew from seven people to three, a cut that NASA's science advisers say would hit research hard.

Dotcom disaster 'good news for biotechnology'

Washington The dotcom crisis seems to have boosted investment in US biotechnology projects. Although overall venture-capital funding dropped by 61% in the second quarter of this year compared with the same period in 2000, the fraction ploughed into life-sciences companies tripled to 13.8%.

Henry Grabowski, an economist at Duke

University in Durham, North Carolina, who specializes in the health industry, says:

"When dotcoms fell out of favour, people began taking another look at biotech. Maybe there is a lot of risk, but once you get a drug approved, you have the possibility of earnings that are fairly steady."

The figures were contained in a joint survey by the National Venture Capital Association, a trade association, and Venture Economics, a private research company.

German agency set to scrutinize Nazi links

Munich Germany's main research-funding agency, the Deutsche Forschungsgemeinschaft (DFG), is to take another look at its own involvement with Nazism before and during the Second World War.

The DFG has already conducted one study into its past. That account, which was published in 1999, was criticized for downplaying the organization's association with the Nazis, particularly its involvement in medical war crimes (see *Nature* **398**, 274; 1999 and **402**, 461; 1999).

A new group of independent historians will now reinvestigate the agency's history between 1920 and 1970, with emphasis on the Nazi years between 1933 and 1945.

The DFG evolved from an earlier funding



Downturn: cosmonaut Sergei Krikalev aboard the troubled International Space Station.

AP

agency established in 1920. Like many other German institutions, it was dissolved after the war, but was re-established in 1949.

It has existed in its present form — as a state-funded but self-governing research agency — since 1951.

Charity pours \$5 million into undersea project

San Diego The Neptune project, an ambitious plan to monitor the microbial life that thrives around undersea fault lines, has received a US\$5-million grant from the W. M. Keck Foundation, a charitable organization based in Los Angeles.

The project aims to create a network of 3,200 kilometres of electro-optical cable to power and communicate with an array of instruments on the ocean floor. The array would cover the 200,000 square kilometres of the Juan de Fuca Plate, a tectonic plate off the western coast of North America. The instruments will form a network of underwater observatories to study the organisms that live around the plate's edges.

The entire project is expected to cost \$250 million. Advocates of the plan had hoped that the National Science Foundation would pay for most of it, but the science agency has not yet decided to do so.

► <http://www.neptune.washington.edu>

Hungry hunters dine on 'extinct' bird

London Half a century is a long time to wait for your favourite food. So when three Indonesian hunters stumbled across a species of game bird not sighted since 1938 and thought to have become extinct, they had no doubts about what to do: they ate it.

The unlucky Bruijn's brush-turkey (*Aepyodius bruijnii*) was caught on the island of Waigeo, off northwest New Guinea. Researchers from Holland and Indonesia identified the bird from its remains, which were salvaged by a local man who knew that they were searching for it.

The finding will, however, help efforts to track down and protect remaining birds in the difficult terrain.

"Once you've got clues that the bird is there you tend to redouble your efforts," says Keith Howman, president of the World Pheasant Association in Reading, Britain. "So while this is bad luck on this particular bird, it is good luck for the species."

Levi-Montalcini appointed Italian senator for life

Rome Rita Levi-Montalcini, who won the 1986 Nobel Prize in medicine for her work on nerve growth factor, was last week appointed



Crowning glory: Levi-Montalcini receives her latest honour from President Ciampi.

senator for life by Italian President Carlo Azeglio Ciampi. This is a rare honour, as the Italian constitution allows the president to maintain a list of only five life senators. Ciampi selected Levi-Montalcini for her "scientific and social merit".

Levi-Montalcini, who is Jewish, was forced into hiding during the Second World War. She left Italy in 1947 to take up a position at Washington University in St Louis, where she conducted the work that earned her the Nobel prize.

Still sprightly at the age of 92, Levi-Montalcini told journalists last week that the honour was "more important than the Nobel prize because it came from my country".

Down and out in Murray Hill

The name Bell Labs is a byword for technological creativity. But its future is now clouded by the financial woes of its parent company, Lucent Technologies. Irwin Goodwin reports.

Of all the world's industrial research centres, Bell Laboratories wears the crown. Bell Labs has been an icon of ingenuity ever since its launch in 1925 by American Telephone & Telegraph (AT&T). At their base in Murray Hill, New Jersey, Bell Labs' researchers pioneered the development of transistors, lasers, optics, digital data transmission, satellite communications and the UNIX computer operating system. Over the years, the labs' work has been honoured with an astonishing six Nobel prizes.

For anyone fascinated by the interface between fundamental physics and high technology, therefore, Bell Labs' present circumstances are a cause for serious concern. The current economic slump has hit all high-tech firms hard. But a series of disastrous business decisions, and a high-profile failed merger, have placed Bell Labs' parent, Lucent Technologies, in especially dire straits. Late last month, recording net losses of \$3.2 billion for the second quarter of 2001, the company announced a restructuring that will see 15,000–20,000 jobs cut from its



current worldwide workforce of 76,000 — in addition to the 19,000 positions shed since the start of the year.

Bell Labs' managers are confident that its scientists, mathematicians, engineers and technicians — some 3,000 of whom are engaged in research, rather than product development — will be protected from the worst of the looming cuts. But as everyone associated with Lucent is forced to tighten their belts, morale is slipping, particularly among researchers who have never known the harsh reality of life in a struggling company. "What is happening here at Bell Labs is an entirely new experience for many young scientists," says one veteran of the labs.

Practically perfect

Speak to some of Bell Labs' best-known alumni and the sense of tragedy becomes all the more tangible. Steve Chu of Stanford University in California, who shared the 1997 Nobel Prize in Physics for developing techniques to cool and trap atoms using lasers, joined the labs in 1978. "We felt like

the Chosen Ones," he recalls. "The joy and excitement of doing science permeated the halls, and the cramped labs and offices forced us to interact with each other. Management supplied us with funding, shielded us from extraneous bureaucracy and urged us not to be satisfied with doing merely 'good' science. Life at Bell Labs was practically perfect."

In this idyllic environment, research flourished — and despite Lucent's financial plight, Bell Labs' scientists are still at the forefront of technological innovation. Current projects include the development of 'electronic paper', portable sheets that can be 'reprinted' at the touch of a button to display different documents (J. A. Rogers *et al. Proc. Natl Acad. Sci. USA* **98**, 4835–4840; 2001). And in March, a Bell Labs team reported the discovery of superconductivity in a plastic material (J. H. Schön *et al. Nature* **410**, 189–192; 2001).

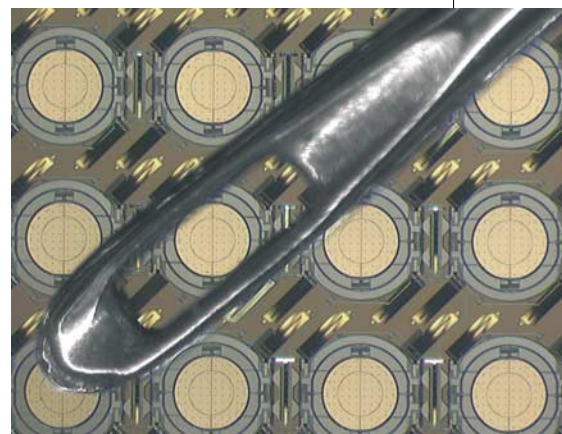
So what went wrong? Many researchers say that the current uncertainty can be traced back to the mid-1990s, when a government



Transistor inventors: William Shockley (seated), Walter Brattain (right) and John Bardeen.



Bell Nobel: Horst Störmer, one of the labs' laureates, was optimistic when Lucent took over.



Needle's eye: these Bell Labs' micromirrors route data by reflecting light between optical fibres.

AP (LEFT AND RIGHT); AP (CENTRE)

antitrust investigation into AT&T was in full swing. "In 1995, the year AT&T was being pummelled in the press and the courts as a monopoly, the mood was about as black as can be," says Jim Eisenstein, a condensed-matter physicist at the California Institute of Technology in Pasadena, who was at Bell Labs for six years in the early 1990s. "We didn't know if we would be working there in a month or two."

But when AT&T was ordered to spin off its systems and technology division in 1996, and the majority of Bell Labs' activities came under the new company, Lucent Technologies, many researchers were optimistic. Horst Störmer spent 23 years at the labs, the last nine as director of physical research, until leaving in 1999 to teach at Columbia University in New York. "Under AT&T, we were considered a bunch of nerds, who might or might not come up with products. We thought Lucent would be technologically driven and run by techies," says Störmer, who shared the 1998 physics Nobel for discovering that electrons acting together in strong magnetic fields can form 'quasi-particles' with charges that are fractions of a single electron's. "Our job was to demonstrate that the company could be scientifically vibrant."

Dimmed outlook

But life under Lucent turned out not to be so rosy. The company took over Bell Labs' research activity on a much smaller turnover than that enjoyed by AT&T — which in hindsight was always going to be difficult to sustain. Industry analysts add that Lucent failed to reverse AT&T's tendency to be slow in turning Bell Labs' bright ideas into marketable technologies. In the late 1990s, for instance, Lucent decided not to push the development of a new generation of optical networking equipment. To the dismay of Bell Labs' scientists, Canada's Nortel Networks, Lucent's principal rival in this field, pressed ahead with the technology, and today its share of the optical networking market is several times that of Lucent.

"This industry is about great products," says Vinod Khosla of Kleiner, Perkins, Caufield & Byers, a venture-capital firm in Menlo Park, California. "And for a long time, Lucent has been coasting on older products."

But this alone does not explain Lucent's current difficulties. Even Nortel is now struggling, as the world's optical networks have more capacity than is needed. Lucent's particular problems stem in large part from a strategy that ran spectacularly into the buffers when the high-tech stockmarket bubble burst last year.

"Things went bad when the managers bet on the wrong horses," says Eisenstein. Lucent had concentrated on providing infrastructure for the new Internet economy. The company had also aggressively acquired about a dozen start-up dotcoms, and loaned money,



Signs of the times: trading in Lucent shares at the New York Stock Exchange in April.

products and services to its customers.

When the crash came, Lucent was disastrously exposed — as is illustrated by two recent losses. The company was building a fibre-optics network for One.Tel, an Australian media firm that collapsed in May. Lucent claims to have lost at least \$488 million on this deal alone. Winstar Communications, another major customer, filed for bankruptcy protection in April, owing Lucent some \$700 million.

The failure in May of a proposed merger with the French telecoms company Alcatel deepened the gloom and further diminished Lucent's sliding stock-market valuation. Now Lucent is restructuring in a bid to rescue its finances. The company's fibre-optics operations have already been sold for \$2.75 billion. Lucent is spinning off its voice and data network business into a company called Ayava and its activities in microelectronics into a firm called Agere. These companies are taking with them only a small percentage of Bell Labs' fundamental research scientists — but their loss will be felt. "We will lose some of the synergism that enabled people to bounce ideas off scientists in other fields," laments one Bell Labs researcher.

No surrender

Senior managers are putting a brave face on the current situation. "We have every intention of staying in the game and continuing to contribute world-class research," says Bell Labs' vice-president, William Brinkman. He points to the high calibre of scientists such as John Rogers, director of condensed-matter physics, who is heading the electronic-paper project, and applied mathematician Wim Sweldens, head of scientific computing research.

"I'm impressed with those joining us now," adds Brinkman. "They're not the sort who want to strike it rich with stock options. They want to publish their discoveries. They want the opportunity to mix with engineers, mathematicians and the variety of physical

scientists here to develop new devices."

But as Lucent's problems deepen, few young scientists are joining Bell Labs. Last year, the labs hired 20 newly graduated physicists and mathematicians, but so far this year only a handful of postdocs have been taken on. Many Bell Labs scientists, meanwhile, are quietly making it known that they are in the market for alternative employment. "We have been enriched by recruiting some of the very best and brightest," says Hans Coufal, manager of science and technology research at IBM's Almaden Research Center in San Jose, California.

When *Nature* visited Bell Labs' headquarters just days before the announcement of Lucent's latest financial results, Brinkman was confident that the spirit of innovation at the labs would help Lucent bounce back. Although Bell Labs' total workforce might be reduced by 25%, he anticipated no forced redundancies in basic research.

But a tour of the labs showed that cost-cutting is rampant. Fluorescent lights have been removed from hallways, staff have been asked to hand back cellphones, and those working late can no longer order pizza on Lucent's tab. A memo circulated a few weeks ago summed up the situation: "If every employee spent \$5 less per month on telephone calls at work, we could save the company up to \$4 million annually."

Phones and pizza are not the real issues, of course — the main concern is whether the labs' tradition of allowing researchers the freedom to explore the areas they find interesting can survive.

Nevertheless, many staff remain loyally optimistic. "I believe these hard times will force us to come up with even better ideas," says mathematician Jelena Kovacevic. "This place will be preserved, I'm sure." Physicists and aficionados of high technology are crossing their fingers and hoping she's right. ■

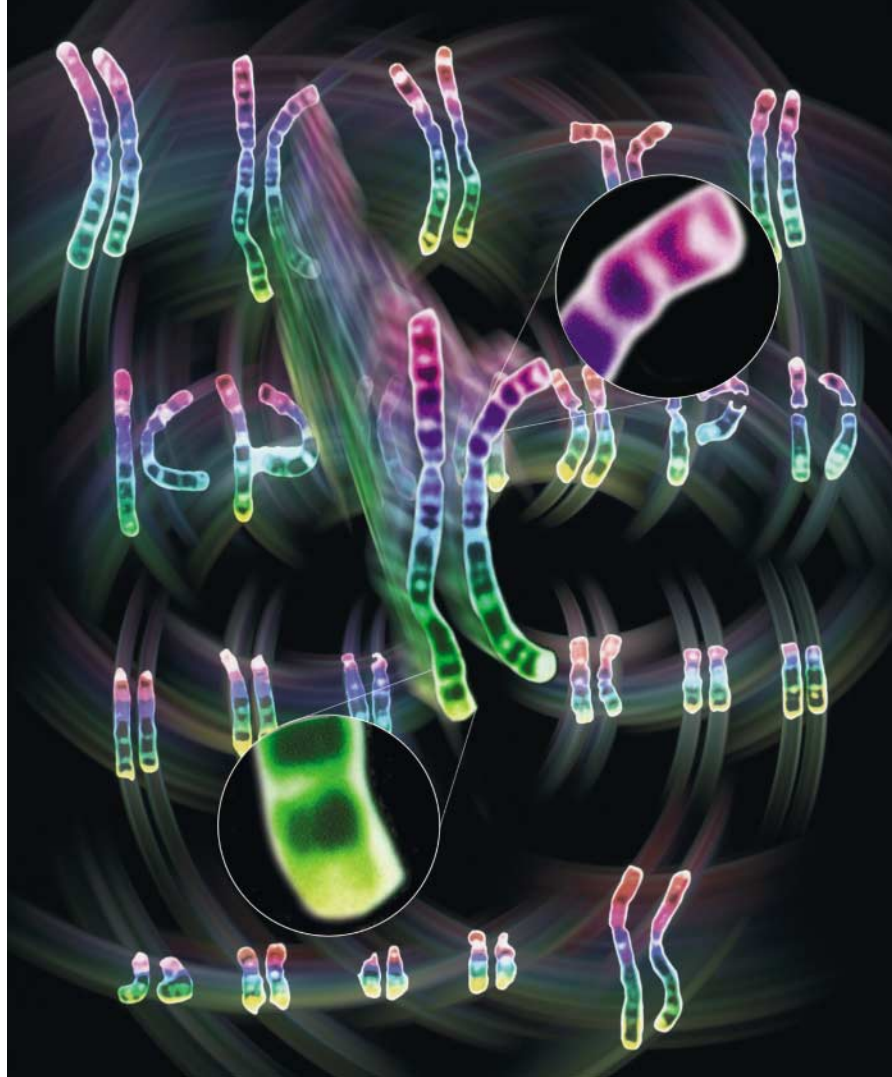
Irwin Goodwin is a freelance writer in Washington.

► <http://www.bell-labs.com>

► <http://www.lucent.com>

Faster, better, cheaper genotyping

Scanning the genome for subtle genetic variations across thousands of individuals may help researchers find genes that underpin susceptibility to common diseases. Marina Chicurel considers the technological requirements.



CNRI/SPL

Back to bases: mapping SNPs to locations on chromosomes could help to unravel disease susceptibility.

The publication in February of draft sequences of the human genome^{1,2} dominated the news worldwide. But for many of the researchers hunting the genes that underlie conditions such as heart disease and cancer, just as important were the less-trumpeted accompanying announcements on the discovery of single nucleotide polymorphisms (SNPs) — points in the genome at which the genetic code can vary by a single ‘letter’.

SNPs — pronounced ‘snips’ — account for most of the genetic variability across human populations. Because they are simple, abundant and widely dispersed, they make excellent landmarks for navigating the genome. As genetic variants that lie close to each other on a chromosome tend to be inherited together down the generations, monitoring SNPs may help gene hunters to trace sequences associated with the susceptibility to common diseases. Doctors might also in the future routinely test for particular SNPs and so tailor drug treatments to each patient’s individual genetic make-up.

In their February paper, researchers at Celera Genomics of Rockville, Maryland, announced the location of 2.1 million SNPs². The International SNP Map Working Group, a coalition of academic labs backed by leading companies and Britain’s Well-

come Trust, published a map containing 1.4 million (ref. 3). The search continues: in some countries, including Japan and China, efforts are under way to identify SNPs specific to their respective populations.

The hunt is on

But mapping SNPs is merely the first step in the hunt for genes involved in disease susceptibility. Researchers must then identify which SNPs are most valuable as markers — many show insufficient variability within a given population, and some are found in repetitive regions of the genome and so do not make useful landmarks⁴. Then comes the task of screening for the useful SNPs in large numbers of people to look for those variations that are associated with particular traits, such as susceptibility to coronary heart disease. And this, at present, is where the available technology is falling short.

Genome-wide gene hunts could require the analysis of hundreds of thousands of SNPs from tens, or even hundreds of thousands of individuals. That sends the number of individual SNPs to be genotyped into the billions⁵. Many researchers are focusing on ‘candidate’ genes already suspected of being linked to a particular trait. But even these more limited efforts can require screening tens of thousands of SNPs in thousands of

individuals. To make such studies possible, the throughput of the world’s SNP genotyping labs must increase by one or two orders of magnitude, and costs will need to be brought down at least tenfold. “The ideal assay will be very quick, cheap and easy,” says Pui-Yan Kwok, an expert on SNP discovery and genotyping at Washington University in St Louis, Missouri. “It is not available.”

Ever since the first large-scale attempts at SNP genotyping started three years ago, dozens of alternative techniques have emerged⁶. “But if you look at them, they’re based on a very few experimental concepts,” says Anthony Brookes of the Center for Genomics Research and Bioinformatics at the Karolinska Institute in Stockholm.

These concepts can be divided into three main categories: reactions, detection systems and formats. Reactions are designed to generate specific molecules based on the presence or absence of a particular SNP. The detection systems are coupled to the reactions to reveal these products. And the formats are the conditions under which the reaction and detection steps take place.

One approach under the reaction category is hybridization, first used on a large scale in 1998 by Eric Lander and his colleagues at the Massachusetts Institute of Technology’s Whitehead Institute for Biomedical

Research⁷. Hybridization depends on the pairing of 'complementary' letters in the genetic code, in which adenines (A) bind to thymidines (T), and guanines (G) to cytosines (C). Lander's team used short synthetic DNA sequences complementary to known SNPs. The sequences were immobilized onto glass 'chips', which were then exposed to a chemically tagged sample of an individual's DNA. The researchers looked for the presence of 500 different SNPs simultaneously by detecting where on the chips each sample hybridized. The chemical tags bound to a fluorescent dye, allowing the chips to be scanned using an optical read-out system.

But hybridization can be difficult — it often needs careful calibration to give reliable results. So many researchers are instead using DNA-manipulating enzymes to reveal the presence of particular SNPs. "Enzymes are highly discriminating," says Scott White, a geneticist at the Los Alamos National Laboratory in New Mexico. They also tend to work reliably without the need for extensive optimization of the experimental set-up.

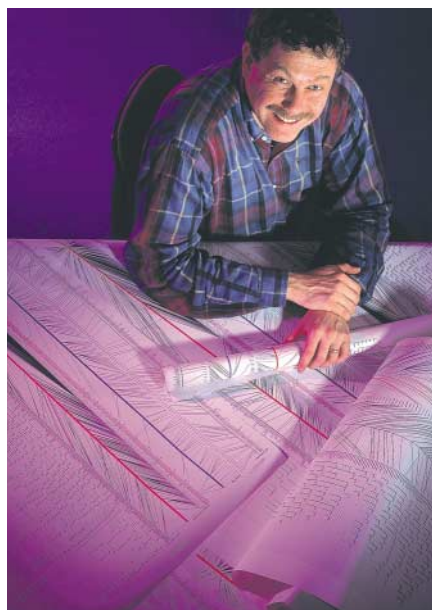
Prime target

Based on these advantages, researchers have developed SNP assays using enzymes that synthesize, cleave or splice DNA. One popular approach, called primer extension, uses a DNA polymerase enzyme to add individual letters of the genetic code, or nucleotides, to a small piece of synthetic DNA called a primer. The primer is designed to hybridize to sequences immediately adjacent to a particular SNP. Once it is in place, the DNA polymerase reads along the rest of the sequence, building a complementary strand of DNA. Researchers can then identify whether a SNP variant is present by monitoring which nucleotide the polymerase incorporates, or fails to incorporate, as it reads along the DNA sample.

Another assay, called Invader and marketed by Third Wave Technologies in Madison, Wisconsin, relies on an enzyme that cleaves DNA⁸. The assay uses two synthetic pieces of DNA, or probes, designed to hybridize to sequences adjacent to a particular SNP. The probes flank the SNP and overlap precisely at the SNP site. If a particular SNP is not present, the overlapping structure will not form. By adding an enzyme that cleaves DNA only when it encounters such overlaps, researchers can assess whether or not the given SNP is present.

The various approaches can be mixed with different detection systems. In primer extension, the DNA polymerase can be fed fluorescently labelled nucleotides, where each of the four nucleotides produces light of a different colour. Alternatively, the extended primer's mass can be measured using mass spectrometry, which can distinguish between DNA molecules differing by only one nucleotide.

Each reaction and detection technique



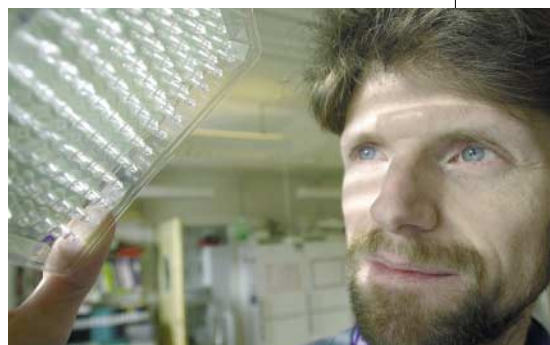
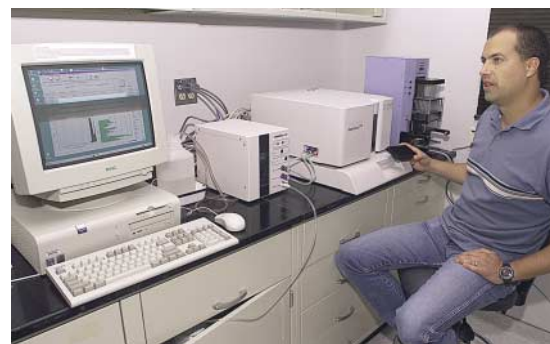
The detectives: Eric Lander (above), Scott White (top right) and Anthony Brookes are all interested in improving methods for identifying SNPs.

has its pros and cons. Many researchers working on large-scale SNP genotyping prefer primer extension because it is robust and flexible. It requires few synthetic DNAs, the design of the primers is simple, and similar reaction conditions can be used for many different primers.

On the detection front, mass spectrometry is popular because it is reliable and yields readily quantifiable results that can be scored easily and rapidly by automated computer systems. Sequenom of San Diego, for instance, markets a technology based on primer extension and mass spectrometry. "What most impressed me was how accurately we could genotype with these out-of-the-box assays," says Kenneth Buetow of the National Cancer Institute (NCI) in Bethesda, Maryland, who is developing methods to streamline SNP genotyping.

Mass spectrometry has the added advantage of not depending on fluorescent labels, which can be expensive. But despite its popularity, the technique has its limitations. Until recently, for example, researchers had to spend a lot of time separating their primer extension products from chemical buffers, the sample DNA, as well as removing the DNA polymerase enzyme and the free nucleotides left over from the reaction. This is because

The throughput of the world's SNP genotyping labs must increase by one or two orders of magnitude.



mass spectrometry requires pure products.

Methods such as the one developed by a team led by Ivo Gut of the French National Centre for Genotyping in Evry, near Paris, have helped circumvent this problem. Gut has boosted the sensitivity of detection by adding a chemical group that modifies the charge on the extended primer⁹. Thanks to this increased sensitivity, simple dilution to lower the concentration of leftover reagents will still give a detectable signal.

But the biggest problem with mass spectrometry is that it generally only allows researchers to screen for up to a dozen SNPs at a time¹⁰. Chip-based hybridization approaches, meanwhile, have advanced to the point at which thousands of SNPs can be screened in parallel. Unless 'multiplex' techniques can be developed for mass spectrometry, argues Michael Boyce-Jacino of Orchid BioSciences in Princeton, New Jersey, the technique ultimately will "hit the wall". Orchid is also marketing a system that relies on primer extension, but offers its clients a variety of detection systems.

Just as detection systems can be mixed and matched with different reactions, the situations under which the reactions occur — the format — can also be varied. When fluorescent tagging is used as a detection system in primer-extension genotyping, many SNPs can be analysed in parallel if the DNA primers for different SNPs are immobilized on a chip. The light given off from each complementary strand built by the DNA polymerase enzyme can then be detected independently. But, compared with assays in which the reagents and products float free in solution, such methods are less flexible. Adding new SNPs to the analysis means that the chips must be redesigned. And the heating and cooling

► required for the primer-extension reaction are difficult to achieve on solid surfaces.

Hoping to get the best of both worlds, some researchers, such as John Nolan and Hong Cai, working with White at Los Alamos, are turning to tiny glass beads about 5 micrometres in diameter¹¹. The researchers first perform standard primer-extension reactions with fluorescently labelled nucleotides in solution.

Extended play

In solution-based assays, it is usually only possible to study one SNP at a time. But by using beads to capture and sort the products of their reactions, the Los Alamos team can study many SNPs in parallel. The researchers place dozens of primers specific for different SNPs in a tube with DNA samples and fluorescently tagged nucleotides. At the end of the primer-extension reaction, the tube contains a complex mixture of labelled products. Then the team adds colour-coded beads carrying 'address tags', pieces of DNA that are complementary to portions of the different primers used in the reaction. As a result, all the products built from one type of primer get attached to beads of the same colour.

The beads can be sorted and analysed using a machine called a flow cytometer. This funnels the sample of beads through a very narrow opening to create a stream in which the beads travel in single file. The cytometer has a laser and a light detector facing the stream, so it can detect the fluorescent colour of each bead as it goes by, as well as the colour of its associated fluorescent nucleotides. It can do so extremely quickly — scoring hundreds to thousands of beads per second. "I think that's going to be one of the concepts that takes us to the next generation of methods," says Brookes.

But Brookes and most other researchers suspect that further advances will be needed to achieve the desired breakthroughs in cost

One of the key bottlenecks to high-throughput genotyping of SNPs is DNA amplification.

and speed. "I think we're a long way away from mature technologies," says Mark Lathrop, director of the French National Centre for Genotyping.

One of the key bottlenecks is the amplification of DNA. Most current assays include a step that produces many copies of a short segment of the sample DNA spanning each target SNP. This amplification is usually necessary because only small amounts of DNA can be harvested from typical clinical samples. Also, the amplification improves the signal-to-noise ratio of the assay, increasing the reliability of detection.

Most genotyping techniques accomplish this amplification using molecular biology's workhorse, the polymerase chain reaction, or PCR. Although PCR is very competent at its job, it is expensive. In addition, setting up PCR to amplify more than 10 targets in parallel is extremely difficult. Those researchers who have achieved multiplex PCR have had to work hard to optimize their systems^{12,13}.

That is why researchers working on SNP genotyping are watching the progress of a team led by Yusuke Nakamura of the University of Tokyo's Human Genome Center. Nakamura is working with sealed cards in which samples are subjected to 100 parallel PCRs, and claims his team can genotype almost 400,000 SNPs a day¹⁴. He says the key lies in the design of the PCR primers, the artificial DNA sequences that define the stretch of sample DNA to be amplified. But some experts remain sceptical. "I don't know how they can do it," says Kwok. A paper outlining Nakamura's methods will appear shortly¹⁵.

Pool cues

Exploring the flip side of multiplex PCR, some researchers are amplifying and genotyping single SNPs from many individuals at once. Working with researchers at Sequenom, the NCI's Buetow has pooled DNA samples from close to 100 individuals and assessed the presence of thousands of SNPs collectively¹⁶. Although pooling obscures the presence of rare SNPs and results in the loss of information on how SNPs are arranged on individuals' chromosomes, it speeds up genotyping immensely. It can, for example, allow rapid comparisons of SNPs from a group of individuals suffering from a particular type of cancer with those who are cancer-free.

Predicting the future of SNP genotyping



Bright idea: Solexa will use laser optics to speed-read an individual's genome.

technology is not easy. The field is moving rapidly, with new approaches springing up all the time. Among the most ambitious ideas being mooted is a novel DNA sequencing technology from Solexa, a British company based in Saffron Walden, near Cambridge¹⁷.

Solexa aims to make chips that will contain up to a hundred million immobilized fragments of single-stranded DNA. The chips will be sequentially washed with solutions containing a single type of nucleotide, each bearing a fluorescent tag, in the presence of a DNA polymerase enzyme, which will try to build complementary strands of DNA. After each wash, lasers will be used to record where the tagged nucleotides have been added, before the tags are chemically removed and the process repeated with a different nucleotide. In this way, claims Solexa, it will be possible to speed-read an individual's genome, SNPs and all, in a matter of days without recourse to PCR.

Whether Solexa's technology will provide what researchers working on SNP genotyping are looking for remains to be seen. But most feel that a technique that is similarly ambitious in its scope will probably be required. "The winner may not even be in the race yet," says Buetow.

Marina Chicurel is a writer in Santa Cruz, California.

1. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
3. The International SNP Map Working Group *Nature* **409**, 928–933 (2001).
4. Marth, G. *et al.* *Nature Genet.* **27**, 371–372 (2001).
5. Abbott, A. *Nature* **406**, 340–342 (2000).
6. Kwok, P.-Y. *Annu. Rev. Genomics Hum. Genet.* (in the press).
7. Wang, D. G. *et al.* *Science* **280**, 1077–1082 (1998).
8. Lyamichev, V. *et al.* *Nature Biotech.* **17**, 292–296 (1999).
9. Sauer, S. *et al.* *Nucleic Acids Res.* **28**, E13 (2000).
10. Ross, P., Hall, L., Smirnov, I. & Haff, L. *Nature Biotechnol.* **16**, 1347–1351 (1998).
11. Cai, H. *et al.* *Genomics* **66**, 135–143 (2000).
12. Fan, J. B. *et al.* *Genome Res.* **10**, 853–860 (2000).
13. Lindblad-Toh, K. *et al.* *Nature Biotechnol.* **18**, 1001–1005 (2000).
14. Cyranoski, D. *Nature* **410**, 1013 (2001).
15. Ohnishi, Y. *et al.* *J. Hum. Genet.* (in the press).
16. Buetow, K. H. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 581–584 (2001).
17. Adam, D. *Nature* **411**, 402 (2001).



Fast-track: Yusuke Nakamura claims his system can genotype nearly 400,000 SNPs a day.

Finding the right questions to ask about the lives of human clones

Child-development experts may have useful information.

Sir — I welcome Lee Silver's call in his Words essay "What are clones?" (*Nature* **412**, 21; 2001) for an informed public debate on human reproductive cloning, but I question his proposed basis for the discussion. Silver concludes that a person produced by nuclear transfer would be "a unique and unpredictable child who had the same DNA sequence as someone else, but nothing more".

I disagree with his implication that a clone would necessarily have the same opportunity for individual development as a child produced by sexual reproduction. The reasons most commonly suggested for producing a clone are to overcome infertility or to replace a dead child. In the first case, the clone would be produced from one of the parents; in the second it would be from a child lost in an accident or after illness. The clone would be physically very similar to the original and have quite a similar personality, because of their shared inheritance. There would be greater similarity to the original in both regards than to any other person except an identical

twin born at the same time as the original.

It seems inevitable that this unusual similarity and the reasons for the production of the clone would influence relationships formed by the child throughout its lifetime. If the original was a dead child in the same family, there is no doubt that the parents wish "to use cloning to bring dead children back to life", as noted by Silver. What then would be the effect — not only on parents, but also on relatives, friends, school teachers and other children — of expectations that the clone would grow up like the original? If a parent were the original, would they have unusual and unreasonable expectations as to how the clone should develop? As the parent aged, how would the cloned child then react to seeing its physical future?

It is concern over these issues that makes me and many others reject the suggestion of cloning a person. The views of those who have studied child development would be very welcome.

Ian Wilmut

Roslin Institute, Roslin, Midlothian EH25 9PS, UK

Why are Indian journals' impact factors so low?

Sir — Despite several limitations, impact factors — produced by the Institute for Scientific Information (ISI) — remain the most widely used, globally acceptable tools to evaluate the quality of journals and research publications. We have looked at the impact factors of Indian journals and find that, for 1999, only 47 journals figured in the ISI's list, all with impact factors of less than 0.6. Of the 5,500 journals from other countries in the ISI's list, 2,286 have impact factors greater than one, including 44 with impact factors greater than 10 and 20 with impact factors greater than 17.

For a country with more than a billion people, a large infrastructure for science and technology, and plenty of scientists, this picture of journal quality is dismal.

Is the coverage of journals for developing countries by the Science Citation Index (SCI) adequate? Is the ISI's monopoly contributing to the problem by restricting coverage or introducing an regional bias, for example between developed and developing countries? The criteria for inclusion in SCI are not known. We have made enquiries to the Indian

National Scientific Documentation Centre (INSDOC), to the National Institute of Science, Technology and Development Studies, to editors of prominent biomedical Indian journals and to other organizations. These have revealed that no Indian agency is involved in analysing these issues at present, though INSDOC has plans to do so.

The response to the questionnaire we sent to these organizations and journals (with a few exceptions, including INSDOC and the *Indian Journal of Medical Research*) was poor, indicating indifference. Most researchers publish their high-quality research in foreign journals with high impact factors, which exacerbates the problem. But they cannot be blamed for this. Administrators use impact factors in making assessments for promotion, recognition, honours and awards. Most official forms for job or grant applications have separate columns for the number of papers published in national and in international journals. An outstanding piece of research published in a less well-known journal might go unnoticed, depriving the author of due recognition.

Is the quality of our publications as poor as it seems? Are impact factors giving a true picture? The reasons for the situation in India must be properly

investigated and remedial measures sought. If scientists, journal editors and learned societies take the initiative in calling for such investigation, this objective will be achieved more quickly.

S. B. Vohora*, Divya Vohora†

*Department of Toxicology, Faculty of Science,

†Department of Pharmacology, Faculty of Pharmacy,

Hamdard University, New Delhi 110 062, India

Genome helpdesk site keeps information public

Sir — The European Molecular Biology Laboratory (EMBL), with additional support from the UK Medical Research Council and the Wellcome Trust, has established a new genome helpdesk, <http://www.ensembl.org>, at the European Bioinformatics Institute (EBI). The EBI is the primary provider of public genome-sequence data within Europe. The US National Institutes of Health has expressed its strong support for the EBI helpdesk, which will complement the existing service, <http://www.ncbi.nlm.nih.gov/genome/guide>, at the US National Center for Biotechnology Information (NCBI).

Together, these initiatives will ensure that the vast potential of the publicly funded genome-sequence databases is fully exploited and freely available for all to use.

The NCBI helpdesk answers more than 300 queries each day from scientists and is an invaluable guide for navigation of the publicly available genome databases. Together with the new EBI site, users will have easy access to an unsurpassed collection of genome sequences and tools for their interpretation. Both helpdesks are staffed by expert teams and rapidly answer queries by e-mail as a public service, available without restriction. It adds to a range of resources provided freely by the EBI and NCBI for commercial and academic scientists to maximize the potential of the public genome databases. These resources are continuously being refined and improved as new genome data are added.

Sir George Radda

Medical Research Council, 20 Park Crescent, London W1B 1AL, UK

Other signatories to this letter:

Mike Dexter Wellcome Trust, London, UK

Fotis Kafatos EMBL, Heidelberg, Germany

Christian Bréchet INSERM, Paris, France

Philippe Kourilsky Institut Pasteur, Paris, France

Joachim Seelig Biozentrum, University of Basel, Switzerland

Alan Bernstein Coalition For Biomedical & Health Research, Ottawa, Canada

Motoya Katsuki National Institute for Basic Biology, Okazaki, Japan

W. G. van Aken ZON NW (Council for Medical Research and Health Development), The Hague, Holland

E.-L. Winnacker DFG, Bonn, Germany

Heidi Diggelmann National Science Foundation, Zurich, Switzerland

Society talks back

The time has come for science to accept that it must leave its cloistered cell.

Re-thinking Science: Knowledge and the Public in an Age of Uncertainty

by Helga Nowotny, Peter Scott & Michael Gibbons

Polity Press: 2001. 278 pp. £50 (hbk), £14.99 (pbk)

Jean-Jacques Salomon

The topic of science as a social institution and its relationship with society has not been covered in such an original fashion since the first seminal papers by Robert Merton in the 1930s and Thomas Kuhn in the 1960s. This book goes far beyond *The New Production of Knowledge* (Sage, 1994), the earlier collection of essays by Michael Gibbons, Helga Nowotny and others.

That book launched the debate on the trend towards a new regime for the production of knowledge and the practice of research. The editors contrasted two regimes. The first, known as mode 1, is the traditional framework for scientific research and its values, which will not disappear. It is deeply entrenched in the universities, where access to 'mainstream' scientific competence will remain based on disciplinary training that is largely insulated from the demands of society.

The second regime, mode 2, represents a new research system, which brings together competences and training from numerous areas. It is multidisciplinary rather than mono- or transdisciplinary, is carried out in non-hierarchical organized groups, and not primarily in institutions such as universities. It is mainly concerned with applications and broad societal problems (from health to environment), and involves an enlarged circle of participants — including the general public through 'citizens' and non-governmental organizations — and has a widened definition of research. The key to mode-2 research is the contextualization of science — the development of knowledge within a particular social context — a process that undoubtedly started in the nineteenth century, in the chemical and electrical industries in particular. The fact that today it involves a different configuration of actors, institutions and norms from the traditional mode means that it is, in effect, a true discontinuity.

Rethinking Science revisits these themes in the form of a single brilliant essay in social theory. It extends the influential reflections of American sociologist Daniel Bell on post-modern social structures and their evolution (see *The Coming of Post-Industrial Society*, Heinemann, 1973; and *The Cultural Contradictions of Capitalism*, Heinemann, 1976). Present-day societies are characterized not

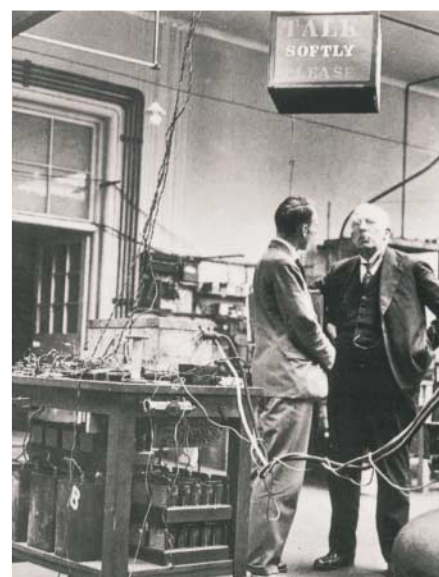


Social pressures: the world of yesterday's scientists must yield to today's priorities.

only by pluralism and diversity, but also by volatility and transgressivity (in the sense of individuals, organizations and cultures acting beyond their traditional boundaries). The co-evolution of science and society has led to increased complexity, unpredictability and irregularity in both spheres. Post-modern society has both a new perception of uncertainty and new means of dealing with risks.

The extent to which society adopts and disseminates new products and practices — from the virtual world of information technology to embryo manipulation — does not only raise new issues of social and cultural order. It also has ethical dimensions that can catch lawyers, politicians and philosophers unawares and undermine existing national, cultural and institutional hierarchies. Breakthroughs and innovations require that choices be made by society in a context of uncertainty and risk. Nowotny *et al.* underline how, in the 1970s, chaos theory provided the wider public with a metaphor for how relationships can be subject to ever-changing patterns of unpredictability. I would have thought, however, that public perception of the social implications of uncertainty went back even further, to quantum theory and Heisenberg's uncertainty principle, which were popularized after the Second World War.

In a mode-2 society, the conceptual and organizational categories of the modern world — state, market, culture, industry, science, educational institutions — have ceased to be recognizably distinct domains, and old distinctions between the 'internal' and the 'external' are becoming problematic. Such conditions allow society to talk back. This reverse communication is so radically transforming science, with its 'private world'



and its own practices, that even the knowledge-based roots of science are being invaded by the forces of contextualization. A very small part remains that is still devoted to context-free science and curiosity-driven research. But the largest part of the enterprise now consists of associations of private, public and industrial laboratories, university researchers and outsiders as well. This is such an enlargement of science's scope, and such an enrichment of its potential in terms of applications, that most of science is no longer seen as, nor even acts as, a discrete cultural domain.

This description raises many questions, and will challenge the traditional scientist's view of his or her autonomous status and independent practice. And why not? It also tends to emphasize the uncoupling of the factual knowledge gained by science from its cognitive authority — a contentious point that mode-1 champions will question at length. In effect, this means that society will not necessarily accept an innovation — say, genetically modified food — just because scientists show that it is safe, any more than

AP

PROF. PETER FOWLER/SPL

politicians will act because scientists show how climate change is a real threat to the planet and mankind.

In the continuing battle between historians and sociologists of science, what is at stake is the issue of the epistemological, or knowledge-based, foundations of science. In this regard, the authors make one of their most radical statements: "The epistemological core is either empty — or, alternatively and perhaps more accurately, crowded with other epistemologies." The controversy may not disappear, but the authors make their point when they give examples — such as Japan and the developing countries — of where Western science, in order to function, has to associate with and come to terms with cultural heritages that define another kind of epistemological approach.

Equally important will be the question of how reliable is the knowledge produced under mode 2, which is so radically different from the peer system of the closely knit community of mode 1. In *Real Science: What it is and What it Means* (Cambridge University Press, 2000), John Ziman considers that the traditional academic ethos, in particular the commitment to disinterestedness, is jeopardized, if not alienated, by the closer links between industry and academia. Socio-economic power, he and others claim, threatens to be the final authority. I would hardly disagree on this point. Suffice to mention the uneasy relations between the commercial genome-sequencing company Celera and the international public consortium for the Human Genome Project, which are all about patents and the prospects for new start-up companies in genetics and involve intervention by bankers, politicians and governments.

Yet Nowotny and her colleagues convincingly address such criticism: after all, reliable knowledge has always been reliable only within limits, and the shift away from weakly contextualized to strongly contextualized knowledge — for instance, from theoretical physics to bioengineering — does not eliminate the need for reliability. On the contrary, it adds robustness, for research, while receiving input from society, is still being tested, verified and validated. The authors introduce a new (but old) notion: they propose that the contextualization of knowledge is taking place in a public arena that they call the 'agora' — the public market-place of classical Greece. To Nowotny *et al.*, this is a place where social knowledge and social participation constrain scientists to reconstruct their authority, to understand and communicate better how science is perceived and to determine how the public trust in science — beyond market and political pressures — can be maintained.

To these authors, the rethinking of science is not "science re-thought", but a new conception — indeed, a new era — of research practice, image and values. This

very exciting book, suffused with a deep general culture and a precise knowledge of the field, ends with 17 points to be debated for the agora, or "how to live with mode-2 society". It should stimulate the interest not only of scientists or specialists in science studies, but also of social scientists, intellectuals and the lay public.

One issue the authors choose not to develop is the influence of the scientific-military-industrial complex in industrialized countries, as if it had no impact on the agora itself. It seems to me, however, that the scientists involved in defence research, who are acting both as "warriors" and as "victims" — as the physicist Freeman Dyson described himself — are the best illustration of how mode 2 combines professional behaviour, the excitement of research, patriotic commitment, corruption, but also scruples, ethical concerns and even attention to public anxiety and criticism.

But this does not prevent the book from being a splendid vision of a probable future world, in which science and society will increasingly overlap and be exposed to the growing expertise and contesting forces of the agora. ■

Jean-Jacques Salomon is honorary Professor of Technology and Society at the Conservatoire National des Arts et Métiers, 2 rue Conté, 75003 Paris, France.

The case of the missing *carpaccio*

The Evolution Explosion: How Humans Cause Rapid Evolutionary Change

by Stephen R. Palumbi
W. W. Norton: 2001. 277 pp. \$24.95

Jerry A. Coyne

Many students open *The Origin of Species* expecting intellectual fireworks, but are disappointed to find a soporific discussion of sheep and pigeons. Yet Darwin's tedious opening chapter on artificial selection was a stroke of rhetorical genius: by recounting the familiar triumphs of animal and plant breeders, he paved the way for his infinitely more heretical evolutionary ideas. (This strategy nearly misfired: one of the publisher's readers, noting that "everybody is interested in pigeons", recommended that Darwin drop the messy stuff about evolution and concentrate instead on birds.)

Although darwinism is now firmly established in scientific and intellectual life, it is far from entrenched in the public consciousness. This is due partly to the prevalence of creationism, but also to the notion — familiar to anyone who teaches premedical students — that evolution has nothing to



do with everyday life. In *The Evolution Explosion*, Stephen Palumbi, a biologist at Harvard University, tries to dispel this view, using Darwin's own strategy of appealing to the reader's experience and hoping that familiarity breeds consent.

Palumbi concentrates on cases in which humans have produced rapid evolution in other species by changing their environments: his examples include the evolution of antibiotic resistance in bacteria, herbicide resistance in plants, pesticide resistance in insects, and changes in the growth rate of fish caused by overfishing. Remarkably, many people familiar with these phenomena have failed to see that they demonstrate evolution driven by selection. There is, for example, a public misconception that 'drug resistance' involves not evolutionary change in pathogenic bacteria, but some process whereby a person becomes acclimated to antibiotics.

Palumbi writes enthusiastically and clearly, and his stories are based on extensive research documented in an appendix. His chapter on AIDS is particularly useful, describing in detail how HIV evolves to avoid the double depredations of our immune system and new generations of antiviral drugs. Elsewhere, we learn that antibiotic resistance in pathogenic bacteria has been caused both by the overprescription of drugs (leading, for example, to the resurgence of drug-resistant tuberculosis) and by the use of antibiotics that increase growth rates in farm animals (producing chickens harbouring dangerous, drug-resistant salmonellae and explaining the absence of *carpaccio di pollo* in Italian restaurants).

Sadly, our understanding of these evolutionary responses seems to have contributed



Evolutionary aid: the widespread use of antibiotics has furthered the evolution of resistant bacteria.

little to solving the attendant medical and economic problems. Combating drug or pesticide resistance usually involves applying more drugs or poisons—solutions that hardly require a sophisticated understanding of evolution. Moreover, new treatments are eventually stymied by further evolution, so that human ingenuity seems nearly impotent in the face of recurring mutation and selection.

By compiling and explaining these diverse cases of anthropogenic evolution, Palumbi has made a useful contribution to the public understanding of science. However, *The Evolution Explosion* suffers from a few problems. Palumbi's narrative runs out of steam in a final chapter about memes, the 'units of culture' (slogans, ideas and inventions, for example) that serve as analogues to genes in theories of social evolution. Despite much attention, 'memetics' has ultimately proved a sterile metaphor, of little value in understanding history or society. It is not clear why Palumbi chose to drag a perfectly straightforward book about science into this pseudo-philosophical quagmire.

The book's style poses an equally serious problem. Straining mightily to achieve what the dust-jacket calls "popular imagery", Palumbi produces an incessant stream of exuberant metaphors and similes that can distract rather than enlighten. In places the prose causes near-physical pain, as in the discussion of memes: "Bad ideas, rejected like anchovy daiquiris, live on only in a few people with fishy breath. Good ideas duplicate quickly and spread far and wide, generating clutches of mental ducklings, with some subsequently turning into brilliant swans and others fated to remain only brain geese."

Finally, although the science is generally accurate, Palumbi's discussion is occasionally confusing or incorrect. For example, he repeats as truth the common belief that artificial selection has made domestic turkeys so dim-witted that during storms they look up at the rain, forget to look down, and drown. This is, in fact, an agro-urban myth that has

been branded an "unfounded turkey rumor" by *Turkey Call*, the official organ of the National Wild Turkey Federation. (Turkeys have, however, suffered greatly from domestication. Responding to human fondness for breast meat, farmers have bred birds too buxom to bonk, and new turkeys must be produced by artificial insemination.)

Nevertheless, *The Evolution Explosion* should help quash the eternal student complaint that evolution is irrelevant. Alas, it is unlikely to change the minds of creationists and advocates of 'intelligent design'. Many who reject darwinism on religious grounds already accept anthropogenic evolution as 'adaptation within a species', but argue that such small changes cannot explain the evolution of new groups of plants and animals. This argument defies common sense. When, after a Christmas visit, we watch grandma leave on the train to Miami, we assume that the rest of her journey will be an extrapolation of that first quarter-mile. A creationist unwilling to extrapolate from micro- to macroevolution is as irrational as an observer who assumes that, after grandma's train disappears around the bend, it is seized by divine forces and instantly transported to Florida. Those not besotted by the anchovy daiquiris of creationism, however, will be convinced by Palumbi's book that evolution is alive and well, if not always welcome. ■

Jerry A. Coyne is in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.

An energetic view of nature

Cosmic Evolution: The Rise of Complexity in Nature

by Eric J. Chaisson

Harvard University Press: 2001. 320 pp.

\$27.95, £19.50

George Ellis

There is a tension in science between, on the one hand, working on the details of particular issues confronting us, immersed in the comfort of our specific technical expertise, and on the other hand stepping back to confront the broader scheme of things and asking how our particular scientific endeavour relates to the 'big questions' facing humanity. Scientists are, by and large, reluctant to address these questions directly.

Eric Chaisson, a senior astronomer and distinguished science educator, has had the courage to tackle such issues. In *Cosmic Evolution* he examines the central question of why we are here, taking as the background the present understanding of our cosmological context. His intention is "to sketch a grand evolutionary synthesis that would

better enable us to understand who we are, whence we came, and how we fit into the overall scheme of things". He focuses on the origins of structure and the spontaneous rise of complexity in nature, in particular, biological life and human intelligence.

Chaisson analyses these issues from a traditional scientific position, under the broad rubric of "Cosmic Evolution ... an underlying ubiquitous pattern penetrating the fabric of all the natural sciences", claiming to provide "a unified scenario of the cosmos, including ourselves as sentient beings, based on the time-honoured concept of change".

Other intrepid souls have trodden similar territory, so what makes this particular synthesis different? Those previous efforts varied in the breadth of context and the variety of themes they addressed, and Chaisson does well on that score. His discourse covers a wide range, from the physics of the early Universe to the origin and nature of life, touching on issues such as the 'anthropic principle' in cosmology, the thermodynamics of non-equilibrium systems, darwinian views on the evolution of life seen in the context of present-day molecular biology, and issues of cultural development. Thus, he takes seriously the modern biological synthesis and also places it in its proper physical and cosmological context, emphasizing interesting causal links. For example, in the end, it is the expansion of the Universe that has allowed complex non-equilibrium structures to emerge. That overall synthesis is welcome.

There is also the choice of a central organizing theme to consider. Other writers' choices have ranged from newtonian determinism, through catastrophe theory, to chaos and complexity theory on the physical side, and neodarwinian evolutionary theory on the biological side. Chaisson's choice is somewhat surprising: although he acknowledges the above approaches, his own scheme centres on energy flows. Thus, he considers in detail the equilibrium and non-equilibrium thermodynamics of many systems, ranging from the early Universe, to stars, to photosynthesis and cell metabolism, and makes this analysis the central feature in a "package of understanding" involving energy flows, physical evolution, natural selection and ordered states. His proposal turns out not to be as misplaced as one might think. Yes, indeed, a thunderstorm has far more energy than any animal—but its volume is vast, so the energy flow per unit volume is small. It is in complex systems such as computer chips and animal brains that it is high.

Chaisson provides impressive evidence that more complex systems are indeed characterized by a higher free-energy flow per unit volume. He gives comparative figures for systems that range from stars to proto-cells to animals to society as a whole, producing a curve of increasing energy-flow

densities with increasing time as the Universe evolved. He relates this curve to the creation of complex structures through the bifurcations characteristic of systems far from equilibrium, together with a generalized principle of natural selection.

Grand schemes stand or fall by their ability to deal with the details of the broad themes they are supposed to synthesize. Although the book covers much material in an enlightening way, I sometimes found it disappointing in this respect. The discussion of the anthropic principle tilts at a straw man that I do not think anyone would seriously espouse. This nice discussion of the important theme of entropy does not explicitly recognize the major unsolved problem of gravitational entropy, although he lays the groundwork to do so. His definition of life is inadequate in that it simply fails to comprehend the full complexity of biochemical systems, and his principle of natural selection is vague and verging on the tautologous.

Finally, how successful is Chaisson in producing the overall integration he intends? The energy-flow issue he focuses on is an important adjunct to the growth of complexity but is not, in my view, the central feature that makes it all possible. High energy-flow density is a requirement, but so are the accumulation of information, for example, and the growth of the ordered structures that make this possible. Indeed, his approach has no real capacity to characterize truly complex systems possessing massive hierarchical ordering, as opposed to less complex but very energetic systems such as the flame of an acetylene torch. The approach might perhaps have been given more substance by relating it to network thermodynamics in complex, hierarchically structured systems, but that has not been attempted here.

Finally, 'grand syntheses' also vary in their degree of ambition, and here Chaisson verges on dangerous territory. His professed aim is to include "all known manifestations of order and complexity in the universe". He wishes, for example, to include cultural evolution in his grand synthesis and to espouse "a new philosophy — a scientific philosophy". But his new philosophy does not touch truly human concerns such as aesthetics, ethics and human culture in a serious way. Nor does it even begin to probe how human cultures deal with the grand themes of life — fear and hope, war and peace, love and death. The grand claims to deal with cultural evolution and to provide a new philosophy are, in the end, not fulfilled — but the journey is interesting and thought-provoking, and the book will serve a useful purpose if it encourages others to think in a synthetic way. ■

George Ellis is in the Department of Mathematics, University of Cape Town, Rondebosch 7700, Cape Town, South Africa.

Science in culture

Volumetric valencies

The molecular visualization tool devised by Preston MacDougall and Christopher Henze.

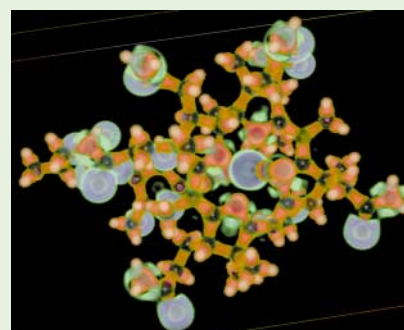
Martin Kemp

Our desire to see what the building-blocks of nature look like seems to be irresistible, even when it is meaningless to talk of their visual appearance in any normal sense. Of all the modern genres of representing the unseeable, none has offered a more beguiling parade of visual delights than the modelling of big molecules. From the polished glyptic formula kits of the nineteenth century to today's proliferating programs of computer modelling, the images have not only satisfied our cognitive urges but have also played key roles in understanding and predicting the properties of molecules that are operating at the very heart of life, disease and death.

As the now quaint-looking ball-and-stick constructions have been superseded by a range of computer programs, such as Per Kraulis's widely used MolScript, so it has become apparent that different modes of representation may highlight quite different kinds of structural information. Accordingly, each system of modelling can handle potentially different aspects of the properties of the molecule. The as-yet-unwritten history of the iconography of molecules would tell of complex symbioses between new kinds of information, new and old representational means, and the research questions that shape the visual grammar of the imagery.

One of the most beguiling of the new systems is that recently devised by the chemist Preston MacDougall, of Middle Tennessee State University, and Christopher Henze, a visualization specialist at the NASA Ames Research Center in California. Their images are based on the topological analysis of the distribution of electron probability density, the cornerstone of Richard Bader's quantum theory of atoms in molecules. Using the laplacian of the probability distribution, their models effectively plot the lumps and hollows, the extrusions and holes in the electron cloud around each nucleus. The resulting sculptures could be described as the joint handiwork of classical and quantum agents.

A fourth dimension is provided by the colour coding, in which a rainbow colour scale is taken as corresponding to a range of charge concentration, from 'cool' blue depletions to 'hot' red concentrations. The transition from green to yellow corresponds to the inflection between depletion and concentration, and white denotes the very highest concentrations. Even with this amount of modelled and coloured information, the static image of an elaborate molecule remains frustratingly complex to unravel. The process of visualization is completed by an animated fly-around facility and by an opacity-function editor that allows



MacDougall and Henze's image of the vitamin B₁₂ complex (cyanocobalamin form) is the joint handiwork of classical and quantum agents.

us to focus and re-focus interactively on features of special concern (see *Theor. Chem. Acc.* **105**, 345; 2001).

The tangibly seductive nature of the tool is not in doubt. But is all the ingenuity worth the effort in scientific terms? The answer is a definite yes, in that it represents an extension of the power of three-dimensional models to predict chemical bonding. Not only do the lumpy configurations reveal very clearly the strong covalent bonding that provides the structural integrity of a molecule, but they are also very effective in denoting sites of noncovalent interactions, the weaker links that typically involve hydrogen atoms at the extremities of large molecules. The plastic modelling tool discloses topologies that permit the identification of reactive sites, including novel ones that are outside the purview of rule-based algorithms. The identification of such sites promises rich potential, most notably in drug design.

As MacDougall himself recognizes, the historical antecedents of such techniques of sculptural 'fitting' go back at least to the 'lock-and-key' model of substrate binding advocated by Emil Fischer at the end of the nineteenth century. There is even an echo of the ideas of Nicholas Lemery, author of *Cours de chimie* in 1675, who was once derided for his conjecture that "chemical combination between two substances, such as an acid and a base, might be accounted for by supposing that the particles of one were sharp, and those of the other porous, and that chemical combination was effected by the fitting of the points into the holes".

There is a nice sense that with the new tool, as is so often the case in scientific visualization, fundamental kinds of visual satisfaction and scientific functionality nourish each other in equal measure. For the animated fly-around, see www.nasa.gov/~chenze/Preston/b12.mpg

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

Avoiding ambiguity

Scientists sometimes use mathematics to give the illusion of certainty.

Sunetra Gupta

What words conceal is as important as what they reveal. Although the essence of raw communication may be clarity, in literature it is the inexact and the imprecise that allow us to push forward the boundaries of human experience and cognition. This is most obvious in poetry, which relies on the flexibility of meaning to record and analyse the breadth and depth of human emotion. For example, the wealth of tenderness in this extract from Seamus Heaney's poem "Sunlight" derives from the mystical alliance between love and a well-worn object:

*And here is love
like a tinsmith's scoop
sunk past its gleam
in the meal-bin*

Yet it also questions our very definition of love. Such poetry highlights not only the ambiguities in the relationships between the words that it uses but can also cause one to pause and reflect upon the relationship between the word and the object or idea to which it refers. And it is not only the reader who is held in this state of productive perplexity, for post-modern literary theory grants the author the prerogative of being equally unaware of the layers of meaning contained within his or her own creation.

The exploitation of ambiguity seems to occupy a much smaller place in the pursuit of scientific knowledge. Notably, the language of mathematics — which has proved to be an indispensable tool in scientific inquiry — distinguishes itself by the lack of ambiguity in its terms. Mathematical metaphors are powerful analytical tools precisely because of the unequivocal relationships between their components, whereas the power of the literary metaphor derives from the incertitude in the connections between its parts. Thus, by their very nature, mathematical metaphors can only be applied to a narrow range of problems: those that lend themselves to reduction into very precise elements, and for which the relationship between these elements can be explicitly declared. Most importantly, this whole artificial exercise has to be able then to comment on some aspect of the problem that would otherwise not have been evident.

But something about the comforting rigidity of the process, its seductive notation, but perhaps mostly its connotations of intellectual privilege, has drawn a diverse selection

of disciplines to the altar of mathematical reasoning. Indeed, the widespread misappropriation of the language of mathematics in the social and biological sciences has to be one of the great tragedies of our time.

Nothing can be sadder than the sight of equations crawling down a page of literary theory, nothing more raucous than the invasion of simple rules of cause and effect into the language of psychoanalysis. Far less obvious in its poverty of reasoning is the inappropriate application of mathematical methods to the analysis of certain scientific problems for which we have no obvious solutions. These projects are usually driven by our inability to cope with the unpredictable — stock-market crashes, hurricanes, earthquakes and epidemics. Although we now have at our disposal some fairly sophisticated methods of characterizing uncertainty, these do not actually enable us to control or even predict the extent of disaster. Used injudiciously in these circumstances, mathematics — and especially mathematical modelling — can serve to obfuscate rather than clarify, or at best add nothing at all to the situation other than the illusion of control.

There are a number of reasons why the language of mathematics may not always provide much insight into a complex reality. At a very simple level, many of the fundamental processes involved, such as consumer choice or movement of livestock, may not be amenable to mathematical formulation. Of greater concern is that, when one is attempting to formalize a set of complicated interactions, assumptions can creep in unawares. This is particularly true when a previously useful mathematical model is retailored to fit a new crisis. It is rather easy in these circumstances to become trapped in, and even comforted by, a prevailing paradigm. It is unfortunate that the assumptions embedded in the mathematical structures employed may not always be obvious to the general public.

There is the danger here that mathematics is being used as a signifier of power much as English is currently used in several post-imperialist cultures. At least its very



Beyond words:
injudicious use of mathematics to analyse events such as stock-market crashes can often confuse rather than clarify.

flexibility sometimes permits English to escape the fate of oppressor's language by mutating into a poetic hybrid, as in some examples of post-colonial literature. Mathematics, however, by virtue of its inflexibility, is liable to be less tolerant of misapplication. No phoenix is likely to arise out of the ashes of a misguided mathematical model.

We are fortunate to have at least two modes of inquiry at our disposal: one that depends upon the fidelity of the word to its referent, and another that conversely makes use of the gulf between a word and its referent, as well as between words themselves. But both may fail, as indeed they have time and time again, in the face of human disaster. It is when a catastrophe occurs that we become acutely aware of the limitations of language, and seek to hide behind a curtain of polemic or an abstruse set of equations. It is in these situations that word becomes completely divorced from its referent, and thus negates both poetic and scientific logic.

The language of mathematical reasoning is no less beautiful for the lack of concealment of meaning. In trying to capture the essence of a system through a minimum of unambiguous relationships between a minimum of unambiguous symbols, scientists and artists are driven by a similar concern for beauty and symmetry, a similar thirst for light. What makes mathematics special is its promise of prophecy, the promise that it will help us understand all mysteries and all knowledge. Without a humble awareness of its limitations, such prophecies can have a very hollow ring.

Sunetra Gupta is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

AP

Design by numbers

R. McNeill Alexander

Optimization theory is a branch of mathematics that was developed largely by economists, and which is used enthusiastically by some biologists and viewed with grave suspicion by others. It seeks the best possible solutions to problems: for example, the best investment strategy for a banker, the best breeding strategy for a bird and the best design for a girder or bone.

The structure and behaviour of organisms are moulded by two powerful optimizing processes: evolution and learning by trial and error. Evolution tends to maximize fitness; that is, roughly speaking, an organism's potential for passing its genes on to future generations. Animals may learn, for example, where to go and what to do to maximize food intake, and how to behave to maximize mating opportunity. The incentive to use optimization theory is not to prove that evolution or learning works, but to check our understanding. If my calculations tell me that a particular pattern of behaviour is the best one possible in given circumstances, and

if real animals do something quite different, then that suggests that I may have failed to understand the issues in hand.

In applying optimization theory, we need to be very clear about what 'best', 'circumstances' and 'issues in hand' mean. A typical problem in optimization would have the following form: choose values for variables x_1 , x_2 , and so on, so as to make some function of x_1 and x_2 as large (or small) as possible. For example, in an analysis of human high and long jumping, I formulated a simple computer model that predicts the heights and lengths of jumps; I then varied run-up speed (x_1) and the angle at which the take-off leg is set down (x_2) and found the combination of x_1 and x_2 that gave the highest or the longest jump.

Often in optimization problems, there are limits to the ranges of values that variables can take. In the case of the jumping problem, there is a limit to the speed at which athletes can run. The model led to the conclusion that long jumpers should run up as fast as possible and set down the take-off leg at a steep angle, and that high jumpers should run up much more slowly and set down the leg at a shallower angle. The predicted speeds and angles agreed well with the speeds and angles that successful athletes actually use. The point of the exercise was not to discover the best way of jumping (it seems best to leave that to the athletes and their coaches), but to check that our understanding of muscle physiology is capable of explaining what athletes actually do.

The suspicious attitude of many biologists to optimization theory is exemplified by one of the anonymous reviewers of my proposal for a book that I am currently writing. He or she complained that my outline emphasized optimization of design, "whereas evolution by natural selection often yields suboptimal but adequate design". A comparison of squid and fish might be used to support this view. Squid swim more slowly than typical fish of similar size, but use more energy in the process. The point that has to be understood here is that evolution is constrained by ancestry. A squid is clearly not the best possible swimmer, but it may be close to the best that can be evolved from a mollusc ancestor. An evolving population may be

Optimization

The structure and behaviour of organisms are moulded by two powerful optimizing processes: evolution and learning by trial and error.

compared to a walker in a mountain range who always walks uphill. Depending on the starting point, this may take the climber to the highest summit, or merely to the top of a foothill. In mathematical language, the squid has failed to reach the global optimum, but it may well be near to a local optimum.

The anonymous reviewer continued: "Optimization criteria may often be multi-fold in character and variable in time, rendering a unitary optimal solution an unlikely outcome." There are two good points here. First, natural selection on squid does not work on swimming performance alone, but on the whole suite of characters that influence fitness, and a change that improves swimming may have a detrimental effect on some other function. Second, environmental changes may move the goalposts. However, trade-offs between benefits and harmful side-effects can be taken into account, and in many cases there seems to be little likelihood of the situation being confused by environmental change. (The requirements for swimming will remain much the same until the sea dries up.) The reviewer concluded, grudgingly, that "this is not to suggest that optimal design does not apply in some cases".

Many biologists' concerns about optimization theory seem to stem from a classic paper by Stephen Jay Gould and Richard Lewontin. These authors pointed out that, because evolution is constrained by ancestry, only local optima may be accessible, as in my example of the squid. They also attacked the uncritical use of an inverse optimization approach (if this animal is the answer, what was the question?). The value to biology of properly applied optimization theory has been splendidly demonstrated by Geoffrey Parker and John Maynard Smith, but their message may have to be repeated many times before the doubters are convinced.

R. McNeill Alexander is at the School of Biology, University of Leeds, Leeds LS2 9JT, UK.

FURTHER READING

Alexander, R. McN. *Optima for Animals* 2nd edn (Princeton Univ. Press, Princeton, New Jersey, 1996).
Gould, S. J. & Lewontin, R. C. *Proc. R. Soc. Lond. B* **205**, 581–598 (1979).
Parker, G. A. & Smith, J. M. *Nature* **348**, 27–33 (1990).
Weibel, E. R., Taylor, C. R. & Bolis, L. (eds) *Principles of Animal Design* (Cambridge Univ. Press, Cambridge, 1998).



Jump to it: theoretical modelling shows that athletes' jumping styles produce optimal results.

A squid is not the best swimmer, but it may be close to the best that can be evolved from a mollusc.

Nifty nanoplankton

Jed A. Fuhrman and Douglas G. Capone

The nitrogen cycle in the oceans may need a rethink. It seems that the ability to transform N_2 gas to a biologically available form may be much more widespread than has been assumed.

All living things need nitrogen. But its most common form, N_2 gas, can be used only by microorganisms that possess the enzyme nitrogenase and can 'fix' nitrogen into a biologically usable form. Fixed nitrogen is a limiting nutrient in much of the ocean, meaning that there isn't enough of it to go round and that in such regions it should be highly advantageous to be a species that can fix nitrogen. The best known and most conspicuous marine nitrogen-fixers are large, photosynthetic cyanobacteria of the genus *Trichodesmium* (see Box 1 and Fig. 1). These are plankton, and occur as filaments up to 0.5 mm long that are sometimes visible as aggregates, but don't seem to be common enough to account for all the nitrogen that we now think is being fixed¹.

On page 635 of this issue², we get an indication as to where part of the answer to this puzzle may lie. Zehr *et al.* report that single-celled cyanobacteria 3–10 μm in diameter (a size referred to as nanoplankton) actively express the nitrogenase gene, and that they may be abundant enough to play a major role in the oceanic nitrogen cycle.

The authors have studied nanoplankton samples from the upper ocean near Hawaii. They show that some constituents of the nanoplankton contain messenger RNA for the *nifH* gene, which encodes part of the nitrogenase enzyme. This is direct evidence that certain cells in this size fraction are expressing *nifH* and so, presumably, are fixing nitrogen. Zehr *et al.* went on to clone and

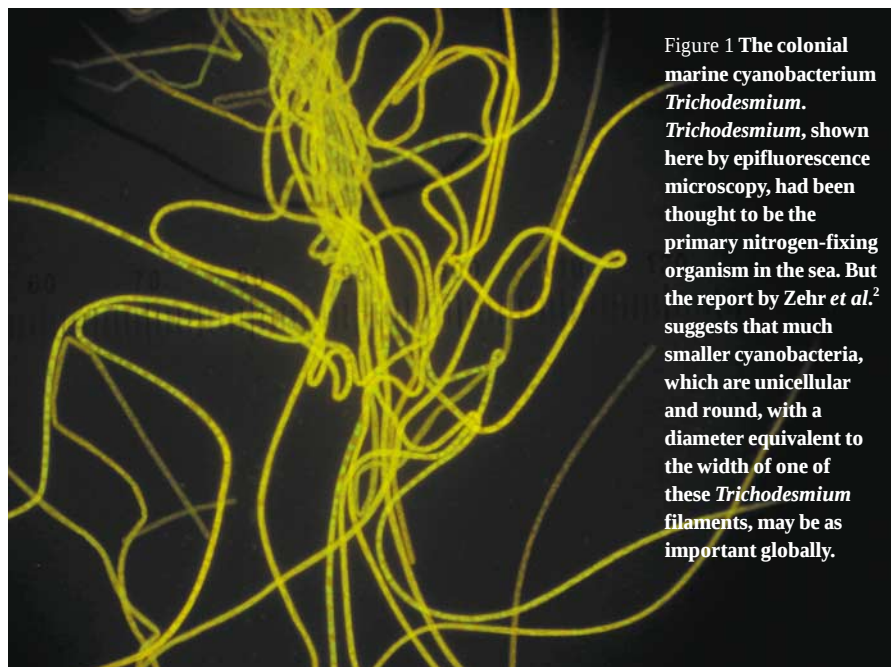


Figure 1 The colonial marine cyanobacterium *Trichodesmium*. *Trichodesmium*, shown here by epifluorescence microscopy, had been thought to be the primary nitrogen-fixing organism in the sea. But the report by Zehr *et al.*² suggests that much smaller cyanobacteria, which are unicellular and round, with a diameter equivalent to the width of one of these *Trichodesmium* filaments, may be as important globally.

sequence the *nifH* messages and compared the sequences with known genes to identify probable source organisms, most of which seem to be unicellular cyanobacteria. Moreover, cyanobacteria 3–5 μm in diameter were cultivated from the samples and were shown to fix nitrogen.

For over a decade, unicellular cyanobacteria 2 μm or smaller in diameter have been known to be important components of the

plankton in all but the polar seas. They reach abundances of 10^8 per litre in warm water, and account for much of the primary production in tropical and subtropical systems³, through the conversion of CO_2 into organic carbon in photosynthesis. But they do not fix nitrogen. Only recently have slightly larger cyanobacteria (3–10 μm) been recognized as potentially significant in the latter respect⁴.

Zehr *et al.*² have now taken this line of

Box 1 The small world of the oceans

Archaea. One of the two groups of prokaryotes, the other being bacteria. Archaea are distinguished from bacteria by numerous genetic and physiological differences. Those members of the archaea that have been cultivated in the lab produce methane or are tolerant of unusually high temperatures or salinities. Uncultivated archaea, such as those now known to be abundant in the deep sea, may have other physiologies. Some archaea fix nitrogen. Evolutionary studies place the archaea, bacteria and eukaryotes as three

fundamentally different domains of life.

Bacteria. The other group of prokaryotes, possessing numerous subdivisions and including most of the cultivated prokaryotic species. Also used as a generic term for organisms that appear to be prokaryotic, but are otherwise unidentified. These are ubiquitous and extremely abundant in all oceans.

Cyanobacteria. The type of bacteria that contain chlorophyll *a* and undergo photosynthesis, generating oxygen. They can be unicellular or

filamentous. This group includes the marine genera *Trichodesmium*, *Prochlorococcus* and numerous other types. Some cyanobacteria fix nitrogen.

Prochlorococcus. Very small cyanobacteria, typically half a micrometre in diameter, that possess the pigments divinyl chlorophyll *a* and *b*. Not thought to fix nitrogen. They are extremely abundant in warm oceans.

Prokaryotes. Cells without membrane-bound nuclei. They include two broadly different groups, the bacteria and archaea.

All known nitrogen-fixing organisms are prokaryotes.

Proteobacteria. A major subdivision of the bacteria that includes a broad array of photosynthetic and non-photosynthetic organisms. Very abundant in sea water. Some members of this group fix nitrogen, including the well-known terrestrial examples that form symbioses with the roots of legumes.

Trichodesmium. Filamentous cyanobacteria, occurring sporadically in warm ocean waters, that fix nitrogen and often form aggregates visible to the naked eye. J. A. F.

investigation a step further. Their research is part of a continuing revolution in the analysis of microbial diversity. Usually, however, previously unknown organisms are found by molecular techniques that detect 'house-keeping genes', such as those for 16S ribosomal RNA, which encode cell components that carry out fundamental processes. Here we see some of the fruits of developing databases of genes that have special functions as well — and, in the case of nitrogen fixation, a function with great biogeochemical significance. Moreover, an intriguing point that Zehr *et al.* mention only briefly is that although cyanobacteria seem to be the source of most of the nitrogen-fixing genes in the nanoplankton samples, 8 of the 27 clones are of another bacterial type, the proteobacteria. This is a broad group that includes mostly heterotrophs (organisms that feed on preformed organic carbon), but also some autotrophs (organisms that use CO₂ as a carbon source). However, new techniques will be needed to quantify the relative contribution of each to nitrogen fixation.

Zehr and colleagues' discovery is far more than a contribution to understanding biodiversity, for it bears on a gap in our biogeochemical balance books. That gap occurs because various lines of evidence imply that there is shortfall in the nitrogen budget of the tropical and subtropical oceans. First, more nitrogen seems to be removed than is added¹. Second, when compared to typical ratios of nitrogen and phosphorus in plankton that don't fix nitrogen^{5,6}, there is often more regenerated nitrogen in deep unlit waters than there is phosphorus. Finally, in systems where nitrogen fixation is relatively unimportant, the nitrogen in biomass is enriched with the heavy isotope, ¹⁵N. In tropical and subtropical seas, however, particulate nitrogen and dissolved nitrate pools in surface waters are depleted in this heavy isotope, and the nitrogen isotopic composition is closer to that of N₂ gas. This is diagnostic of N₂ fixation¹.

Nitrogen fixation by *Trichodesmium* seems insufficient to explain these observations⁵. If nitrogen fixation by nanoplankton turns out to be a globally significant source of nitrogen, we are on our way to getting an explanation for the discrepancies. Zehr *et al.* have made preliminary estimates using isotopic tracer techniques: it looks as though the process may indeed constitute a large item in the marine nitrogen budget, but much more work will be necessary to fill in the actual numbers.

What about possible constraints on nitrogen-fixing marine microorganisms that might diminish their activity? They do not suffer from a lack of nitrogen. But they may be limited by other essential elements, such as phosphorus and iron, which are also usually in short supply. Atmospheric dust is one of the main sources of iron for the

marine environment, and there is a close correspondence between patterns of dust input to the tropical ocean and the distribution of excess nitrate. The implication is that oceanic nitrogen fixation may be largely controlled by iron input^{5,6}, although where iron input is high, phosphorus may become limiting^{7,8}. The same controls may regulate nitrogen fixation in the nanoplankton.

The closer we look at the oceans, the more important the tiniest organisms appear to be. In the 1970s, epifluorescence microscopy revealed abundant, previously unknown communities of unicellular cyanobacteria and bacteria up to 1 µm in diameter, leading to the recognition of their respective roles as important primary producers and consumers of about half the total carbon flow through the marine ecosystem⁹. Later, flow cytometry revealed that a close relative of the cyanobacteria, *Prochlorococcus*, is one of the most abundant algae in the sea^{3,10}, although it had been completely overlooked by older techniques. Electron-microscopic studies then identified marine viruses as remarkably — and unexpectedly — abundant and active¹¹. And with molecular biological approaches, marine archaea have been found to be very abundant in the deep sea, although their ecological significance remains a mystery^{12,13}.

Zehr *et al.* now appear to have uncovered yet another key role for the single-celled marine prokaryotes. Determining the global distribution of cyanobacteria in the nano-

plankton, and seeing whether they routinely fix N₂, will be a challenge. As Zehr *et al.* show, the molecular biological and isotopic-tracer methods are available. We can identify unicellular cyanobacteria through their fluorescence signature and size them by flow cytometry, and can recognize the presence of the nitrogenase enzyme using immunological probes. But although the technological tools are at hand, it will require a large-scale sampling and analytical programme to fully meet the challenge. ■

Jed A. Fuhrman and Douglas G. Capone are in the Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089-0371, USA.

e-mails: fuhrman@usc.edu

capone@usc.edu

1. Capone, D. *Curr. Opin. Microbiol.* **4**, 341–348 (2001).
2. Zehr, J. *et al.* *Nature* **412**, 635–638 (2001).
3. Campbell, L. & Vaulot, D. *Deep-Sea Res.* **40**, 2043–2060 (1993).
4. Zehr, J. P., Carpenter, E. J. & Villareal, T. A. *Trends Microbiol.* **8**, 68–73 (2000).
5. Michaels, A. F. *et al.* *Biogeochemistry* **35**, 181–226 (1996).
6. Gruber, N. & Sarmiento, J. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
7. Wu, J., Sunda, W., Boyle, E. & Karl, D. *Science* **289**, 759–762 (2000).
8. Sañudo-Wilhelmy, S. A. *et al.* *Nature* **411**, 66–69 (2001).
9. Azam, F., Fenchel, T., Gray, J. G., Meyer-Reil, L. A. & Thingstad, T. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
10. Chisholm, S. W. *et al.* *Nature* **334**, 340–343 (1988).
11. Fuhrman, J. *Nature* **399**, 541–548 (1999).
12. Fuhrman, J., McCallum, K. & Davis, A. A. *Nature* **356**, 148–149 (1992).
13. Karner, M., DeLong, E. & Karl, D. *Nature* **409**, 507–510 (2001).

Molecular electronics

Momentous period for nanotubes

David Goldhaber-Gordon and Ilana Goldhaber-Gordon

When conductors are reduced to molecular dimensions they can develop exotic properties. Physicists have now directly confirmed unusual electron behaviour in carbon nanotubes.

Since their discovery¹ in 1991, carbon nanotubes have been touted for their technological promise as molecular wires and junctions². But experiments with nanotubes are also strengthening our understanding of the quantum-mechanical universe: electronic behaviours that were once purely theoretical are now being addressed experimentally. On page 617 of this issue³, Lemay *et al.* advance the state of the art by looking inside a nanotube and studying the electrons within.

A carbon nanotube is a single sheet of graphite rolled up into a tube⁴. It can be thought of as a long and narrow molecule, sometimes tens of micrometres in length, but only a few nanometres in circumference. This narrow circumference is part of what makes nanotubes so interesting. But their length has been an obstacle to quantum-

mechanical studies. When electrons have a large space to fill, their energy levels pack closely together, making it difficult to study individual quantum-mechanical states.

Lemay *et al.* tackled the problem by slicing out a bite-sized piece from a long nanotube. To study this short (34-nm) segment, they used a scanning tunnelling microscope (STM). In the STM technique, a finely sharpened metal tip is scanned just above the surface of a sample. Electrons 'tunnel' across the gap from the tip to the sample, carrying a small but measurable current. In the simplest approach, the voltage from tip to sample is held constant, and the tip is moved up or down to maintain a constant flow of electrons. The tip moves up as it passes over an atom, so a record of tip movement provides a picture of the sample at atomic resolution^{5,6}.

A more sophisticated STM measurement

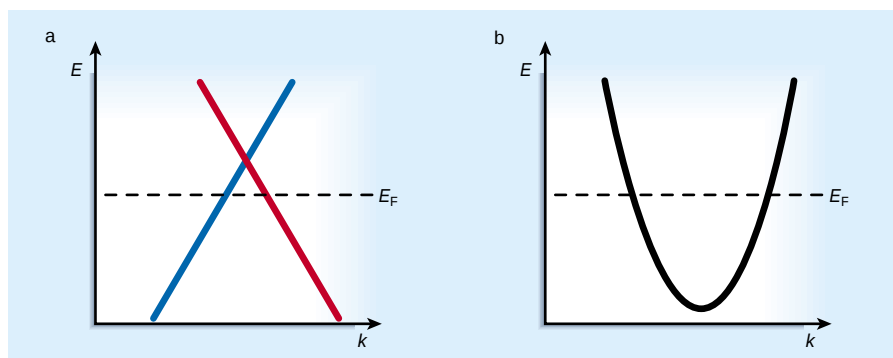


Figure 1 Electron band structures of nanotubes and bulk metals. **a**, In a metallic carbon nanotube, left-moving electrons (red) and right-moving electrons (blue) belong to two different bands with distinct microscopic structures. Within each band, energy, E , is a linear function of momentum, k . Electrons are filled up to the Fermi level, E_F ; this level is near the crossing of the bands, but may vary slightly depending on the environment of the tube. Another band crossing, at very different momentum, is (safely) ignored in our discussion. **b**, In a normal bulk metal, all electrons that contribute to conduction, both left-moving and right-moving, belong to a single parabolic band. In the new work, Lemay *et al.*³ confirm theoretical predictions about nanotube band structure.

allows individual electron wavefunctions to be imaged⁷. A wavefunction, $\psi(x)$, describes the position at which an electron with a given energy, E_ψ , is most likely to be found; the wavefunction squared is the 'probability density'. Areas of high probability density are those in which electrons easily tunnel into or out of the sample. To measure the probability density, the STM tip is again scanned over the sample. This time, as the tip scans the sample, the tip-sample voltage, V , is shifted slightly up and down — intentionally and periodically. The current, I , shifts in step with the voltage, and the ratio of the change in current to that in voltage (dI/dV) can be measured. An especially large oscillation in current indicates that electrons are tunnelling into or out of a quantum state in the sample.

Lemay and colleagues³ performed this dI/dV analysis at three different settings of average tip-sample voltage, each probing a different electron energy in the nanotube. At each energy they scanned across the length and width of the tube, acquiring a distinctive image of spots and lines. The images matched the predicted form⁸ of wavefunctions in nanotubes, showing frequent spatial oscillations with the expected period and orientation (see Fig. 1 of the paper, page 618). These are the first complete experimental images of electron wavefunctions in a nanotube.

More importantly, through detailed analysis of many wavefunction images, Lemay and colleagues have confirmed theoretical predictions⁴ about the 'band structure' of nanotubes. Band structure describes the allowed states of electrons in terms of their energy and momentum: a given band has a unique energy for each momentum. Carbon nanotubes come in two flavours: semiconducting and metallic, depending on how the graphite sheet is wrapped up. Lemay *et al.* studied metallic tubes, which have two bands

that intersect at a single point (Fig. 1a).

Why is the band structure of a nanotube of interest? For one thing, it is unusual. In normal bulk metal, electrons occupy a single, parabola-shaped conduction band (Fig. 1b). In a metallic nanotube, however, two distinct, linear bands cross. At a given energy, one of the linear bands carries left-moving electrons, the other right-moving ones. In addition, one band is constructed from molecular bonding states, the other from antibonding states, so that the wavefunctions of left- and right-moving electrons look very different on an atomic scale. This has an effect on electron behaviour: to switch its direction of movement, an electron must also switch from a bonding to an antibonding state (or vice versa). This restriction suppresses changes in direction, so an electron in a metallic nanotube tends to move persistently in one direction^{9,10}. This situation resembles that of neutrinos, which are elementary particles of (almost) zero mass. Each neutrino's spin depends on its direction of motion¹¹ — changing direction requires a change of spin orientation. In contrast, in a normal metal, left- and right-moving electrons are part of the same parabolic band, have the same microscopic character, and are easily interconverted.

The linear, two-band configuration of metallic nanotubes was predicted⁴ soon after their discovery. Earlier this year, Liang *et al.*¹² provided the first experimental evidence of this configuration by measuring patterns of interference in electron transport through nanotubes. Lemay and colleagues³ have now provided a more direct verification. By analysing their wavefunction images, they detected interference between electrons in the two different bands. As a result of this interference, the net probability density oscillates slowly along the length of the tube. The effect is much like playing the same note simultaneously on two strings of a violin. If



100 YEARS AGO

We have received several papers by Prof. Sommerfeld, dealing with the theory of the diffraction of Röntgen rays. One of these is published in the *Zeitschrift für Mathematik und Physik*, xlv. 1, 2, and abstracts are also given in the *Physikalische Zeitschrift*, ii. The special problem which forms the subject of Prof. Sommerfeld's work is the mathematical investigation of the results of the hypothesis put forward by Wiechert and Stokes, according to which Röntgen rays consist in an impulsive disturbance propagated through the ether. The author considers the problem of diffraction past a screen in the form of a half-plane and allied problems, and compares his results with those found by Haga and Wind and others. The single non-periodic impulse may be said to represent one extreme case of ray-propagation, while the purely periodic wave represents the other extreme. While actual Röntgen rays and light rays probably only approximate to these extreme cases, the agreement between Prof. Sommerfeld's conclusions and experimental results affords considerable evidence in favour of the above theory of Röntgen rays.

From *Nature* 8 August 1901.

50 YEARS AGO

It has generally been assumed that the only carboxylic acid present in the fruit of Bramley's Seedling apple is malic acid; but in 1949, when examining chromatograms run in *n*-butanol-formic-acid-water (40 : 10 : 50 v/v) of methyl alcohol extracts of pulp tissue of young Bramley's Seedling apples, we noted that several spots having an acid reaction to bromophenol blue appeared on the chromatograms. In addition to a relatively large spot corresponding with malic acid ($R_F = 0.49$), there appeared a second well-defined spot ($R_F = 0.18$) and traces of a third spot ($R_F = 0.07$). It was possible to wash some of the acid ($R_F = 0.18$) from the chromatogram and carry out tests on the solution so obtained... It would appear, therefore, that the new acid might well be a dihydroxy tricarballic acid. The acid appeared to decrease in amount as the fruit ripened and also appeared to be present in the pulp of the fruit of peach and plum. One of us has since examined other varieties of English apples for the presence of the new acid. It has been found in much greater quantity in young Worcester Pearmain and young Cox's Orange Pippin apples.

From *Nature* 11 August 1951.

the strings are slightly out of tune with each other, the sound will 'beat' up and down in intensity. The beat frequency is simply the difference between the frequencies of the two notes. Similarly, the spatial beat frequency in these experiments is the difference between the momenta of the electrons in the two bands. The authors measured this momentum difference as a function of energy, confirming the predictions of band-theory calculations.

The existence of two separate bands in metallic nanotubes also has technological implications. An obstacle to using nanotubes as electronic wires is that the conducting electrons sometimes reflect backwards, by bouncing off a structural defect in the tube or off a smoother variation in electrical potential caused by charges outside the tube. In metallic tubes, resistance to switching the direction of motion makes electrons unlikely to reflect off a smoothly varying potential. This protection is not strong enough to prevent reflections from structural defects in a nanotube, but fortunately, improved techniques for growing nanotubes have raised the prospect of entirely defect-free tubes¹². Together, these properties suggest that nanotubes may soon be able to conduct

electrons over many micrometres, making them a viable, much smaller alternative to conventional electronic wires.

David Goldhaber-Gordon, currently of the Harvard Society of Fellows, will from 1 September be in the Department of Physics and the Geballe Laboratory for Advanced Materials, Stanford University, 476 Lomita Mall, Stanford, California 94305-4045, USA. e-mail: davidg@post.harvard.edu

Ilana Goldhaber-Gordon is in the Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

e-mail: ilana@post.harvard.edu

1. Iijima, S. *Nature* **354**, 56–58 (1991).
2. Andriotis, A. N., Menon, M., Srivastava, D. & Chernozatonskii, L. *Phys. Rev. Lett.* **87**, 066802 (2001).
3. Lemay, S. G. et al. *Nature* **412**, 617–620 (2001).
4. Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes* (Academic, New York, 1996).
5. Odom, T. W., Huang, J., Kim, P. & Lieber, C. M. *Nature* **391**, 62–64 (1998).
6. Wildoer, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. *Nature* **391**, 59–62 (1998).
7. Crommie, M. F., Lutz, C. P. & Eigler, D. M. *Nature* **363**, 524–527 (1993).
8. Kane, C. L. & Mele, E. J. *Phys. Rev. B* **59**, R12759–R12762 (1999).
9. Ando, T. & Nakanishi, T. *J. Phys. Sci. Jap.* **67**, 1704–1713 (1998).
10. McEuen, P. L. et al. *Phys. Rev. Lett.* **83**, 5098–5101 (1999).
11. Goldhaber, M., Grodzins, L. & Sunyar, A. W. *Phys. Rev.* **109**, 1015–1017 (1958).
12. Liang, W. et al. *Nature* **411**, 665–669 (2001).

Genome sequencing

The ABC of symbiosis

J. Allan Downie and J. Peter W. Young

The latest bacterial genome to be completely sequenced has three separate parts and as many genes as yeast. The bacterium needs these genes for its complex life in and around its legume plant partner.

It is a truth universally acknowledged, that there are only two kinds of bacteria. One is *Escherichia coli* and the other is not. Anything that *E. coli* does is a universal truth about bacteria; anything it does not do must

be a specialization. This *coli*-centric view has made life a little easier for generations of students, but it has taken some knocks lately. Not only have *E. coli* fans seen their favourite bacterium beaten to the genome finishing

line¹ by several outsiders^{2–5}, but the genome of the laboratory workhorse, *E. coli* strain K12, looks paltry even in comparison with better-endowed *E. coli* strains⁶. As the number of completely sequenced genomes increases week by week⁷, we are beginning to see just how rich life can be in the real bacterial world.

Freshly completed is the sequence of *Sinorhizobium meliloti*^{8–11}. This bacterium is a 'rhizobium' — it can form nodules¹ on the roots of legume plants, converting (fixing) atmospheric nitrogen to a biologically usable form. The *S. meliloti* genome is revealed in four papers by Galibert and colleagues in *Science* and *Proceedings of the National Academy of Sciences*^{8–11}. Why so many papers? One answer is that this organism essentially has three genomes (Fig. 1). Another is that the sequence will reveal wider truths about bacteria.

The *coli*-centric view of a normal bacterial genome is that it consists of a single circular chromosome. There may also be one or two plasmids. These small circular DNA molecules are present in many copies; they carry a few genes that can be very handy (such as those conferring antibiotic resistance) but are generally unnecessary. So when we find a large circular genetic element, twice the size of some whole bacterial genomes, maintained at just one or two copies per cell and carrying more than 1,000 genes, surely we are looking at a chromosome? Not if it is pSymA, at 1.35 million base pairs (Mbp) the smallest of the three elements of the *S. meliloti* genome⁹. This is a megaplasmid, and is not essential for growth.

Next up in size is pSymB, another megaplasmid at 1.7 Mbp, which is essential — for reasons that are now clear¹⁰. For instance, it encodes the cell's only transfer RNA that recognizes the nucleotide triplet CCG, and so is vital for protein synthesis. Finally, *S. meliloti*'s 'real' chromosome

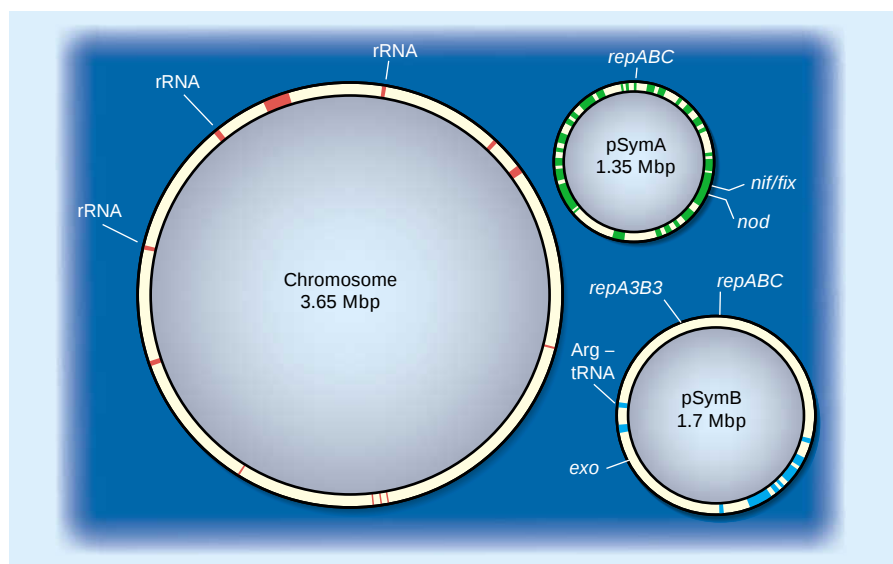


Figure 1 The three components of the *Sinorhizobium meliloti* genome^{8–11}: a chromosome and two megaplasmids. Red, green and blue regions have a guanine + cytosine (G + C) content of less than 60% (averaged over 10-kilobase windows). The positions of some genes are shown, including those needed for the synthesis of ribosomal RNA (rRNA) and for plasmid replication (rep genes), as well as the gene encoding the essential transfer RNA (Arg-tRNA) that recognizes the nucleotide triplet CCG. Also shown are the gene regions required for the bacterium to form nodules on the roots of legumes (nod genes), for the formation of external polysaccharides (exo genes), and for symbiotic nitrogen fixation (nif/fix genes).

weighs in at 3.65 Mbp. It carries all three sets of the genes for making ribosomal RNA, and most of the genes needed for basic metabolism¹¹. The genome of another rhizobium, *Mesorhizobium loti*¹², is on a similar scale, although it has a chromosome of 7.0 Mbp and two plasmids of 0.35 Mbp and 0.2 Mbp.

Clearly we need to rethink our concept of bacterial genome organization; it is no longer straightforward to draw a distinction between plasmids and chromosomes. In their mode of replication, pSymA and pSymB are definitely plasmids: unlike the *S. meliloti* chromosome, they have the replication and stability genes (*repABC*) that are common to most rhizobial plasmids¹³. On the other hand, pSymB, like the chromosome, carries essential genes.

Perhaps the most intriguing difference between the genetic elements of *S. meliloti* is that, as Galibert and colleagues^{8–11} find, pSymA has an average of 60.4% guanosine and cytidine (G + C) nucleotides, whereas the chromosome and pSymB have significantly more (62.7% and 62.4%, respectively). This is partly explained by the higher incidence on pSymA of mobile genetic elements, which tend to have a low G + C content. Yet many other genes on pSymA — including nearly all known nodulation-related (*nod*) genes — also have a much lower G + C content than do typical chromosomal genes (Fig. 1).

Differences in G + C content between accessory genes (needed for specialist functions such as nodulation) and housekeeping (essential) genes are common in bacterial genomes. They are usually interpreted as a sign that the accessory DNA originally came from a bacterium with a different genomic make-up^{6,12}. Although plausible in many cases, this explanation hits a snag with rhizobia. Nearly all known *nod* genes — which are unique to rhizobia — have a low G + C content, yet all rhizobial chromosomes have a high G + C content. This is true even for *Burkholderia* strain STM678, the most distantly related rhizobium yet discovered¹⁴. If the *nod* genes evolved in a genome that matched their low G + C content, we haven't found it yet. Alternatively, is it possible that different parts of the genome could diverge in composition while sharing the same cell, perhaps because of differences in mutation or selection pressures? This is plausible, although the idea is more widely accepted for eukaryotes than for bacteria.

The authors⁹ also find that the genome of *S. meliloti* is large for a bacterium, with over 6,200 genes — as many as yeast. Yet even these do not reflect the full repertoire of genes potentially available to rhizobia. Fewer than half of the 400 or so genes identified on one of the plasmids of a closely related *Sinorhizobium* strain¹⁵ are found in *S. meliloti*⁸, while a third of the 7,000 genes in *M. loti*¹² are missing from *S. meliloti*⁸.

Such large and variable genomes may be necessary for growth and survival in the complex environments of the soil and plant root. Many of the genes on the *S. meliloti* megaplasmids encode proteins that allow the organism to adapt to different environments. About 14% of the genes on pSymB appear to be involved in the production of cell-surface polysaccharides, which are probably important for survival and to allow the bacterium to attach to the surface of plant roots. Another 20% are devoted to solute uptake. On pSymA, 14% of the genes are likely to be involved in importing and exporting molecules, and about 8% are related to nitrogen metabolism. These, together with the many plasmid genes for using diverse substrates, may enable *S. meliloti* to survive in soil, ready to jump at the chance of infecting a legume root.

Rhizobia have specific and intimate interactions with eukaryotic cells: they adhere to plant roots, invade root cells and stimulate cell proliferation. Although the *nod* genes appear to be unique to rhizobia, the *S. meliloti* genome sequence also reveals several possible parallels with the way pathogenic bacteria interact with animal cells. The *S. meliloti* chromosome¹¹ has counterparts of genes that are involved in the invasion of human red blood cells by the Oroya fever bacterium *Bartonella*; in the virulence of

Shigella; in interactions between enteropathogenic *E. coli* and epithelial cells; and in the lysis of blood cells by *Treponema*. Studies of these and other rhizobial genes will illuminate both bacterial pathogenesis and the nitrogen-fixing symbiotic partnership between rhizobia and legumes. ■

J. Allan Downie is in the Department of Genetics, John Innes Institute, Colney Lane, Norwich NR4 7UH, UK.

e-mail: allan.downie@bbsrc.ac.uk

J. Peter W. Young is in the Department of Biology, University of York, York YO10 5YW, UK.

e-mail: jpy1@york.ac.uk

1. Blattner, F. R. *et al. Science* **277**, 1453–1457 (1997).
2. Fleischmann, R. D. *et al. Science* **269**, 496–512 (1995).
3. Fraser, C. M. *et al. Science* **270**, 397–404 (1995).
4. Bult, C. J. *et al. Science* **273**, 1058–1073 (1996).
5. Tomb, J.-F. *et al. Nature* **388**, 539–547 (1997).
6. Perna, N. T. *et al. Nature* **409**, 529–533 (2001).
7. <http://www.tigr.org/tdb/mdb/mdbinprogress.html>
8. Galibert, F. *et al. Science* **293**, 668–672 (2001).
9. Barnett, M. J. *et al. Proc. Natl Acad. Sci. USA* **10.1073/pnas.161294798** (2001).
<http://www.pnas.org/cgi/content/abstract/161294798v1>
10. Finan, T. M. *et al. Proc. Natl Acad. Sci. USA* **10.1073/pnas.161294698** (2001).
<http://www.pnas.org/cgi/content/abstract/161294698v1>
11. Capela, D. *et al. Proc. Natl Acad. Sci. USA* **10.1073/pnas.161294398** (2001).
<http://www.pnas.org/cgi/content/abstract/161294398v1>
12. Kaneko, T. *et al. DNA Res.* **7**, 331–338 (2000).
13. Palmer, K. M., Turner, S. L. & Young, J. P. W. *Plasmid* **44**, 209–219 (2000).
14. Moulin, L., Munive, A., Dreyfus, B. & Boivin-Masson, C. *Nature* **411**, 948–950 (2001).
15. Freiberg, C. *et al. Nature* **387**, 394–401 (1997).

Chemistry

On the threshold of stability

Heinz D. Roth

Carbenes are short-lived compounds containing a highly reactive carbon atom, which makes them difficult to study. A stabilized derivative may lead to new magnetic materials.

Carbon usually has four atoms or groups bonded to it. Many reactions in organic chemistry involve breaking one of these covalent bonds, so chemists are naturally interested in the reactive and unstable species of carbon that fleetingly form during these reactions. The most common carbon intermediates are trivalent species, which form only three bonds. Covalent bonds form when adjacent atoms share two electrons, so many trivalent carbon compounds have one unpaired electron. Intermediates containing a divalent carbon atom that forms only two covalent bonds, such as CH₂, are even more reactive than their trivalent cousins. Such species, called carbenes, have two non-bonding electrons on the same carbon atom. Despite the importance of carbenes to organic synthesis, their short microsecond lifetimes make them difficult to study. On page 626 of this issue, Tomioka and colleagues¹ report the successful preparation of the first triplet carbene with a long lifetime

— almost 20 minutes at room temperature. A truly stable version of this particular carbene could lead to useful magnetic materials.

Studying unstable carbon intermediates requires the reactions to be slowed down or the measurements speeded up. Fast time-resolved spectroscopic methods make it easier to study short-lived intermediates at room temperatures, whereas conventional spectroscopic methods can be used at low temperatures because reaction rates are slower. Alternatively, adding appropriate groups to the reactive carbon can stabilize the transient species. Following on from the discovery by Gomberg in 1900 of the first fairly stable trivalent carbon species, the existence of trivalent compounds was confirmed by stabilizing the elusive species to form isolated molecules, and, later, by fast spectroscopic methods. The chemistry of trivalent carbon is now well understood, but the story of carbenes has taken much longer to unfold. It wasn't until the 1950s

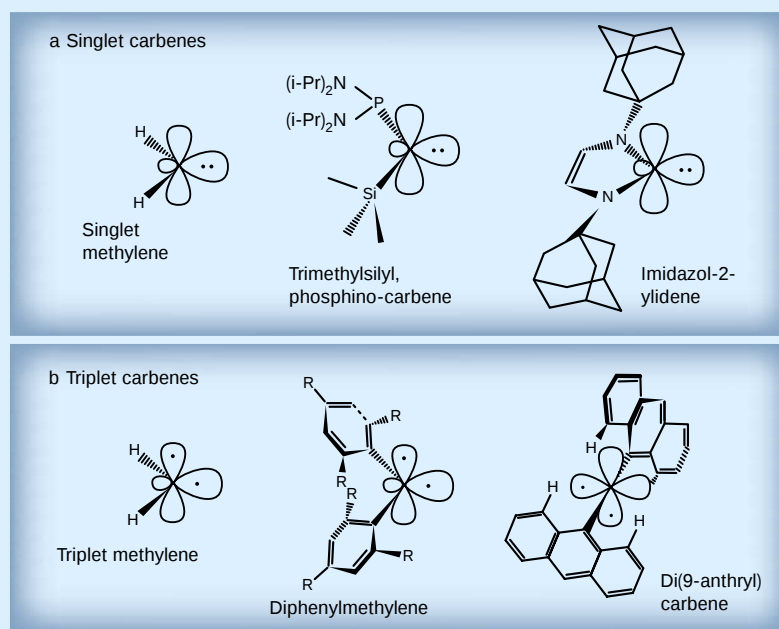


Figure 1 Stabilization of carbenes. Chemical reactions involving a carbon atom in an organic compound sometimes involve breaking two of its four bonds. This results in the formation of highly unstable intermediates known as carbenes, which have two possible electronic states. a, In the singlet state, both non-bonding electrons occupy the same orbital, as in singlet methylene, $^1\text{CH}_2$. Replacing the hydrogen atoms with electronegative groups can stabilize this carbene. So trimethylsilylphosphino-carbene forms a stable liquid³ and imidazol-2-ylidene forms stable crystals⁴. b, Triplet carbenes, such as triplet methylene, $^3\text{CH}_2$, in which the non-bonding electrons occupy different orbitals, are much harder to stabilize. Although di(9-anthryl)carbene is more stable than diphenylmethylene, its lifetime in solution at room temperature is only 0.5 microseconds.

that carbenes were unambiguously identified in chemical reactions, and many attempts to create stable carbenes have failed.

There are two types of carbene species, both of which contain a divalent carbon atom with six electrons — two electrons shy of the stable electron configuration. Four of these electrons are involved in bonding, the other two are non-bonding; they can be paired or unpaired. Typical carbenes have an in-plane σ -orbital (like those in typical single bonds) and a p-orbital perpendicular

to it (like typical double bonds). Singlet carbenes have a pair of electrons in the σ -orbital and an empty p-orbital — for example, the singlet methylene, $^1\text{CH}_2$ (Fig. 1a). Triplet carbenes, on the other hand, have two unpaired electrons, one each in the σ - and p-orbitals (see triplet methylene, $^3\text{CH}_2$; Fig. 1b). The orbital configurations of these carbenes dictate their reactivities, and so are key to attempts to stabilize them. The parent carbene, CH_2 , is highly reactive in either the singlet or triplet state; indeed, CH_2 was once

called the “most indiscriminate reagent known in organic chemistry”². But related carbenes in which one or more of the hydrogen atoms have been replaced by a different atom are much less reactive.

The stabilization of carbenes can involve thermodynamic or kinetic factors. For example, singlet carbenes can be stabilized by electron donation from electronegative substituents because of their vacant p-orbitals, and triplets can be stabilized by sharing the unpaired electron with adjacent atoms, lowering the thermodynamic energy of the species. Alternatively, access to the reactive divalent carbon by other reagents can be kinetically prevented by bulky side groups. Among singlet carbenes, a trimethylsilylphosphino-carbene will form a stable liquid³ and imidazol-2-ylidenes can form crystals⁴ (Fig. 1a). But are these substituted species truly stabilized carbenes or do they lack essential carbene features?

These questions can partly be answered by theoretical calculations of the carbene structures. For example, calculations for trimethylsilylphosphino-carbene revealed a Si–C–P angle of $\sim 138^\circ$, different from the angle in $^1\text{CH}_2$ of $\sim 104^\circ$. It also has a C–P bond (~ 160 picometres) that is too short for a single bond, and a planar arrangement of groups around the phosphorus atom. These features suggest that phosphorus forms a double or triple bond with the divalent carbon, making the compound a phosphonium ylide or a phospho-acetylene, rather than a phosphino-carbene⁵. In contrast, the imidazol-2-ylidenes have an N–C–N angle of $\sim 102^\circ$, typical of $^1\text{CH}_2$, and the divalent carbon carries only a small negative charge. This compound is further stabilized by kinetic factors, as the bulky adamantyl groups shield the divalent carbon from approaching reagents. So stable singlet carbenes are possible, and further variations are to be expected.

In contrast, stable triplet carbenes have been more elusive than their singlet sisters because they have two unpaired electrons in close proximity. Attempts to stabilize triplet carbenes are based on delocalization (sharing) of the unpaired electron(s) (Fig. 1b). For example, the p-electron of diphenylmethylene ($\text{R} = \text{H}$ in Fig. 1b) is delocalized, but its stability is limited and the σ -electron remains localized⁶. An attempt to combine delocalization with kinetic stabilization by bulky substituents in the ortho- and para- positions (say, $\text{R} = \text{CH}_3$ in Fig. 1b) failed to stabilize the corresponding carbene⁷. However, di(9-anthryl)carbene, a carbene with more extended aromatic substituents, is further stabilized by both electrons being delocalized⁸. Even so, its lifetime in solution at room temperature is just 0.5 microseconds.

The stabilized compounds in Fig. 1b provided the starting point for the work of Tomioka's group. They first improved on

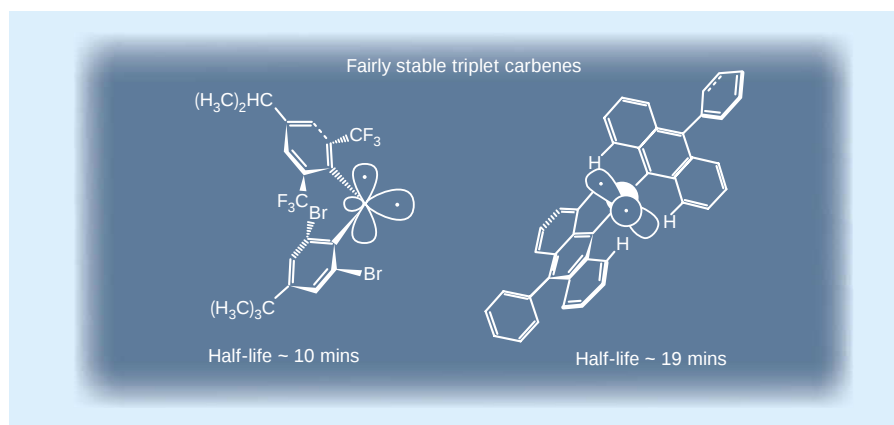


Figure 2 Carbenes that can almost be bottled. In two studies^{1,9}, Tomioka and colleagues have further stabilized triplet carbenes. Left, this triplet carbene is a more stable version of diphenylmethylene (Fig. 1b), with a half-life of 10 minutes at 273 K. Right, by adding phenyl groups to the anthryl groups of di(9-anthryl)carbene (Fig. 1b), Tomioka *et al.*¹ have increased its half-life by eight orders of magnitude to a record 19 minutes at 300 K.

diphenylmethylene by incorporating bulkier ortho- and para- substituents (a combination of CF_3 , $\text{CH}(\text{CH}_3)_2$, Br and $\text{C}(\text{CH}_3)_3$). The resulting carbene (Fig. 2, left) has a half-life of 10 minutes at 273 K; the authors claim it "can almost be bottled"⁹. The significant progress reported in the current paper¹ grew out of the observation that di(9-anthryl) carbene (Fig. 1b) will form a trimer in which three species are coupled together through bond formation between the anthryl groups. To prevent this lifetime-limiting reaction, the reactive positions on the anthryl groups were substituted with phenyl groups (Fig. 2, right). One might expect the additional groups to further delocalize the unpaired electrons and yield an extended trimer. But the phenyl groups are actually rotated out of the plane of the adjacent anthryl groups, thereby efficiently blocking trimer formation. The resulting species has a record half-life for a triplet carbene of 19 minutes at 300 K.

The route to 'fairly stable' triplet carbenes has been more difficult than that for the singlet sisters, but it has its own rewards. Triplet carbenes would make useful building

blocks for organic ferromagnetics, if they were not seriously limited by their instability. The unpaired electrons serve as magnetic spins in organic molecules. The emergence of stable triplet carbenes opens up a new dimension in this research. Still, further improvements pose significant challenges to the ingenuity of future researchers. ■

Heinz D. Roth is in the Department of Chemistry and Chemical Biology, Rutgers University, 610 Taylor Road, New Brunswick, New Jersey 08854-8087, USA.

e-mail: roth@rutchem.rutgers.edu

1. Tomioka, H., Iwamoto, E., Itakura, H. & Hirai, K. *Nature* **412**, 626–628 (2001).
2. Doering, W. v. E., Buttery, R. G., Laughlin, R. G. & Chaudhuri, N. J. *Am. Chem. Soc.* **78**, 3224 (1956).
3. Igau, A., Grützmacher, H., Baceiredo, A. & Bertrand, G. *J. Am. Chem. Soc.* **110**, 6463–6466 (1988).
4. Arduengo, A. J. III, Harlow, R. L. & Kline, M. J. *Am. Chem. Soc.* **113**, 361–363 (1991).
5. Dixon, D. A., Dobbs, K. D., Arduengo, A. J. III & Bertrand, G. *J. Am. Chem. Soc.* **113**, 8782–8785 (1991).
6. Doetschman, D. C. & Hutchison, C. A. Jr *J. Chem. Phys.* **56**, 3964–3982 (1972).
7. Zimmerman, H. E. & Paskovich, D. H. *J. Am. Chem. Soc.* **86**, 2149–2160 (1964).
8. Wasserman, E. et al. *J. Am. Chem. Soc.* **93**, 6335–6337 (1971).
9. Hirai, K. & Tomioka, H. *J. Am. Chem. Soc.* **121**, 10213–10214 (1999).

Signal transduction

Barriers come down

Ulrich Siebenlist

A protein known as erythropoietin might be useful in preventing the death of nerve cells in acute brain injury. But how does it work? Crosstalk between two signalling pathways could be the answer.

Different biological signals can have quite distinct effects on gene regulation, so one might expect them to work through largely separate intermediary molecules within cells. At first glance that certainly seemed to be the case for two common signalling pathways. One of these pathways is switched on by inflammation or stress, and ultimately activates transcription factors — proteins that influence gene expression — from the NF- κ B family. The other is activated by hormones or by growth or regulatory factors, including one known as erythropoietin, and targets a different family of transcription factors, the STATs. Writing on page 641 of this issue, however, Digicaylioglu and Lipton¹ break down the presumed barrier between these pathways. They report that erythropoietin prevents stressed nerve cells from committing suicide by activating NF- κ B proteins. Unexpectedly, it does this through an enzyme called Jak2, which is typically associated with the STAT pathway.

The NF- κ B family of transcription factors is best known for its role in immunity. But these proteins also have other tasks: they are used by many cells when the environment is stressful. In this regard, one of their

main roles is to inhibit the programmed cell death (apoptosis) pathways that can be activated in these circumstances. They do this by inducing the expression of genes whose protein products oppose apoptosis. But it has been less clear which intermediary molecules lead to the activation of NF- κ B under these conditions.

From their studies of how erythropoietin protects neurons from apoptosis, Digicaylioglu and Lipton¹ provide an unexpected answer to this question. This role of erythropoietin has only recently come into focus — it is known mainly for its ability to stimulate the production of red blood cells in bone marrow. Its expression is increased in kidneys whenever low levels of oxygen (hypoxia) occur, a condition sensed by the transcription factor HIF-1. Erythropoietin is used to treat patients for anaemia, after bone-marrow transplants for example.

But erythropoietin is also produced in the brain in response to the harmful chemicals (oxidative or nitrosative stress) that may be generated by inflammation, hypoxia–ischaemia (tissue damage caused by oxygen deprivation), or in neurodegenerative disorders. Moreover, neurons express receptors

that detect erythropoietin. One of its roles in the brain is to stop stressed neurons from dying. Administration of erythropoietin in animals can limit neuronal damage in the brain's cerebral cortex during experimentally induced hypoxia–ischaemia. And, if animals are 'preconditioned' by short-term hypoxia–ischaemia followed by reoxygenation, which results in the production of erythropoietin by the brain, they are protected from later, more prolonged hypoxia–ischaemia². So erythropoietin is under the medical spotlight for a second reason: it might be useful in preventing the loss of nerve cells in some neurodegenerative diseases or after stroke.

But how does erythropoietin work its magic on nerve cells? This is where Digicaylioglu and Lipton¹ come in. They first show that cultured nerve cells from the cerebral cortex that are pretreated with erythropoietin are protected from apoptosis in models of the damage caused by neurodegeneration or hypoxia. Motivated by the knowledge that NF- κ B also protects stressed cells, Digicaylioglu and Lipton find that erythropoietin leads to rapid (within minutes) and sustained activation of this transcription factor in cultured nerve cells. They further show that activated NF- κ B is required for protection by erythropoietin. The authors then identify Jak2 as essential in this process. This enzyme is a protein kinase — it phosphorylates target proteins — and is involved in signalling from the erythropoietin receptor to STAT proteins in non-neuronal cells. Interference with Jak2 activity blocks the protective effects of erythropoietin in neurons¹.

Typically, activation of the I κ B kinase complex is needed before NF- κ B can be switched on (Fig. 1a, overleaf), and this event seems to be required here, too. But an additional, unusual mechanism may contribute as well¹. NF- κ B is normally kept in an inactive state outside the nucleus by its binding partner, a member of the I κ B family (the prototypical member is I κ B α). The I κ B kinases phosphorylate these inhibitors on two specific serine amino acids. This leads to the degradation of the inhibitors and the release of NF- κ B, which can then migrate into the nucleus and activate target genes. An alternative mechanism for freeing NF- κ B involves the phosphorylation of a tyrosine residue in I κ B α ³, although the details and significance remain hazy. Digicaylioglu and Lipton provide evidence that both types of phosphorylation occur when neurons are exposed to erythropoietin, although it remains to be seen whether both mechanisms are required to activate NF- κ B and protect neurons. Finally, Jak2 can directly phosphorylate the tyrosine residue of I κ B α *in vitro*.

I κ B kinases are activated by many signalling pathways, in particular those initiated by molecules involved in inflammation.

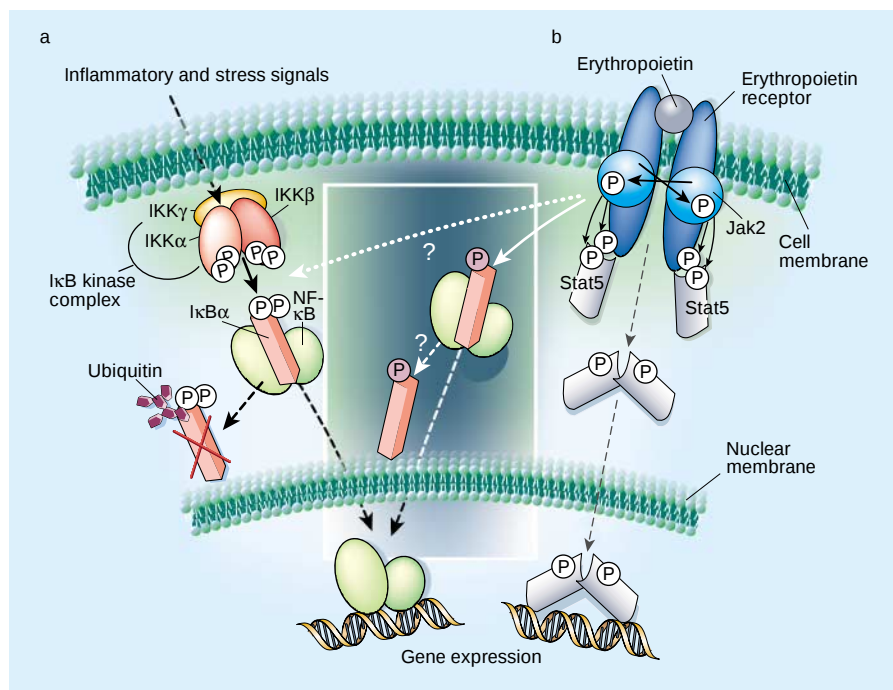


Figure 1 Crosstalk between two seemingly disparate signalling pathways. **a**, The classical NF- κ B pathway, induced by stress or inflammation. This pathway involves the activation of the I κ B kinase complex, the phosphorylation (circled 'P') of I κ B on serine residues, its labelling with a small protein (ubiquitin) and degradation, and the release of NF- κ B, which moves to the nucleus. **b**, The classical pathway that is induced by the growth factor erythropoietin in non-neuronal cells. On binding of erythropoietin to its receptor, the enzyme Jak2 activates itself by phosphorylation. It also phosphorylates the receptor, which recruits the transcription factor Stat5. After Stat5 is also phosphorylated by Jak2, it dimerizes and moves to the nucleus. Centre, the crosstalk between these two pathways that might occur in neurons¹. This crosstalk might involve the typical activation of I κ B kinases, as well as the atypical phosphorylation of I κ B α on a tyrosine residue. Dotted arrows indicate events that are not understood in detail; dashed arrows indicate multistep processes.

Enzymes of the Jak family, on the other hand, are associated with, and mediate signalling from, several receptors involved in the regulation of growth, differentiation and immune functions. When the Jak proteins activate themselves, their next substrates are the associated receptors, which, when phosphorylated, can recruit transcription factors from the STAT family. These in turn become the final substrates of the Jaks (Fig. 1b). Although NF- κ B and STATs (or the products of their gene targets) may collaborate or antagonize each other in gene-specific contexts, the upstream signalling pathways have not previously been shown to communicate — but they are now connected in erythropoietin-stimulated neurons¹ (Fig. 1, centre).

It is not clear why erythropoietin should have this effect only on neuronal cells. Maybe there are neuron-specific linking proteins that allow Jak2 to activate the I κ B kinases and to phosphorylate I κ B α . Further questions include whether other growth factors or hormones that signal through Jak2 can also activate NF- κ B, and whether Stat5 (the better-known target of Jak2) collaborates with NF- κ B to protect neurons. Finally, it might seem difficult to reconcile these results with the fact that the short-lived activation of NF- κ B has actually been associated with

neuronal apoptosis⁴. But what may distinguish this situation from the anti-apoptotic effects described here is that the activation of NF- κ B is sustained after preconditioning with erythropoietin¹.

If the proposed communication between the two pathways is confirmed, and the intermediate molecular events by which Jak2 activates NF- κ B in neurons are worked out, this system might provide a model for how otherwise separate signalling systems can be linked in a potentially cell- and signal-specific manner. We might anticipate the discovery of cell-specific proteins that link Jak2 with its targets. We might also envisage similar scenarios for other, normally unrelated signalling pathways. Such combinatorial diversity would engender specific cells with great flexibility and unexpected biological responses. It would also pose untold challenges for cell biologists. ■

Ulrich Siebenlist is in the Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892-1876, USA.

e-mail: usiebenlist@niaid.nih.gov

1. Digicaylioglu, M. & Lipton, S. A. *Nature* **412**, 641–647 (2001).
2. Siren, A. L. et al. *Proc. Natl Acad. Sci. USA* **98**, 4044–4049 (2001).
3. Livolsi, A. et al. *Eur. J. Biochem.* **268**, 1508–1515 (2001).
4. Lezoualc'h, F. et al. *J. Neurosci.* **18**, 3224–3232 (1998).

Daedalus

Encapsulated gas

Daedalus once invented a 'fractal' concrete. Ordinary concrete contains gravel, the interstices of which are filled with sand, the interstices of which contain reactive cement suspension. Daedalus' concrete had big spaces filled with big particles, the gaps between them filled by smaller ones, the gaps between those filled by smaller ones still, and so on.... The tiny, densely reticulated network left for binding cement could contain high-performance polymer.

Suppose the nested particles were replaced by air bubbles. With each designed to fit neatly in the gaps left by the next larger size, the result would be an immensely complicated foam, nearly all air (or whatever gas was used to blow it). It would be enormously strong for its weight. Evaporation would be poor as a setting reaction. The solvent would take ages to escape through the many almost monomolecular layers, and might take the gas with it. Polymerization would be better; carbonization might be better still. So DREADCO chemists are at work.

With a liquid monomer such as a superglue or a substituted acrylate, the result should be a wonderfully light, translucent foam. Blown with hydrogen or helium, it might be lighter than air, a building-block for Zeppelins. Blown with air, it would rival the aerogels as a thermally insulating filling for two-wall skylights. Indeed, if the bubble-size could be kept away from a wavelength of light, it might even be transparent. As a paint, it should be a wonderful absorber of sound. Weak sounds would be almost completely absorbed by 'pumping' the small bubbles flat at each cycle. Strong sounds should suffer little loss. Thus in a church, ringing calls to repent should survive unattenuated, while the small random noises of the congregation itself would be lost.

The biggest challenge, of course, is carbonization. Battery-makers long for a way of storing hydrogen, possibly under the inherent pressure of its monolayer on benzenoid carbon, using nanotubes or their derivatives. The DREADCO team reckon that a carbonized foam, taken to almost atomic dimensions, should be able to store vast amounts of hydrogen, and deliver power electrolytically via a carbon electrode. It could then be recharged with hydrogen again. With luck both small-scale applications such as mobile phones, and even big ones such as hydrogen-powered cars, could use the new electrode.

David Jones

Pattern of focal γ -bursts in chess players

Grandmasters call on regions of the brain not used so much by less skilled amateurs.

The brain's medial temporal lobe structures are thought to be important for the initial formation of long-term memory^{1,2}, and active memory is indicated by bursts of γ -band activity in these and other areas of the association cortex^{3,4}. Here we use a new technique of magnetic imaging to compare focal bursts of γ -band activity in amateur and professional chess players during matches. We find that this activity is most evident in the medial temporal lobe in amateur players, which is consistent with the interpretation that their mental acuity is focused on analysing unusual new moves during the game. In contrast, highly skilled chess grandmasters have more γ -bursts in the frontal and parietal cortices, indicating that they are retrieving chunks from expert memory by recruiting circuits outside the medial temporal lobe.

The 'chunking' theory of chess playing⁵

suggests that expert memory is based on a large database of chunks in long-term memory. A chess grandmaster studies and practises for at least 10 years to learn more than 100,000 patterns (memory chunks). Consequently, grandmasters can 'recognize' the key elements in a problem situation much more rapidly than amateur players. Experts differ not only in the extent of their knowledge, but also in its organization. High-level processing elements, such as structuring knowledge and planning, assist in accessing the respective chunks⁶.

We tested 20 male players (aged 42 ± 14 years), each with more than 10 years of tournament and training practice. Ten professional grandmasters scored between 2,400 and 2,600 on Elo's chess-skill rating scale⁵; amateur players ranked 1,700 and above. Magnetoencephalographic recordings were made while subjects played against a computer and were scanned in the 5 seconds after each move by the computer program for focal γ -bursts (20–40 Hz; Fig. 1a). Examination of single slices indicates pronounced activity in the region of the perirhinal and entorhinal cortex, hippocampus and related structures in amateur players, but not in grandmasters.

There was a strong negative correlation ($r = -0.84$) between the relative share of dipoles in these structures and Elo chess skill (Fig. 1b). The correlation was also significant when only the 12 players who had lost in the game were included in the analysis (draws and wins were achieved only by grandmasters). There was no relationship between the length or complexity of the game and the Elo score.

These marked differences in the distribution of focal brain activity during chess playing point to differences in the mechanisms of brain processing and functional brain organization between grandmasters and amateurs. Lesions in structures that are activated in amateur players impair recent memory while leaving remote memory intact². Grandmasters seem to rely more on remote than on recent memory.

High-level processing elements⁶ may also have contributed to the observed differences. The chunking theory of memory states that the number and nature of chunks that chess experts can hold in long-term memory can be used to predict chess performance⁶. Our results indicate that the activation of expert memory chunks produces focal γ -band activity in the neocortex, whereas amateur players primarily encode and analyse new information, tasks that activate the medial temporal lobe and the hippocampus. It is possible that these structures play only a transitional role during the establishment of expert memory in the neocortex.

Ognjen Amidzic, Hartmut J. Riehle, Thorsten Fehr, Christian Wienbruch, Thomas Elbert

University of Konstanz, Universitätsstrasse 10, Fach-D30, 78457 Konstanz, Germany
e-mail: ognjen.amidzic@uni-konstanz.de

1. Bontempi, B., Laurent-Demir, C., Destrade, C. & Jaffard, R. *Nature* **400**, 671–674 (1999).
2. Squire, L. R. *Psychol. Rev.* **99**, 195–231 (1992).
3. Pulvermüller, F., Keil, A. & Elbert, T. *Trends Cogn. Sci.* **3**, 250–252 (1999).
4. Singer, W. et al. *Trends Cogn. Sci.* **1**, 252–261 (1997).
5. Elo, A. E. *The Rating of Chess Players, Past and Present* (Arco, New York, 1978).
6. Gobet, F. *Cognition* **66**, 115–152 (1998).

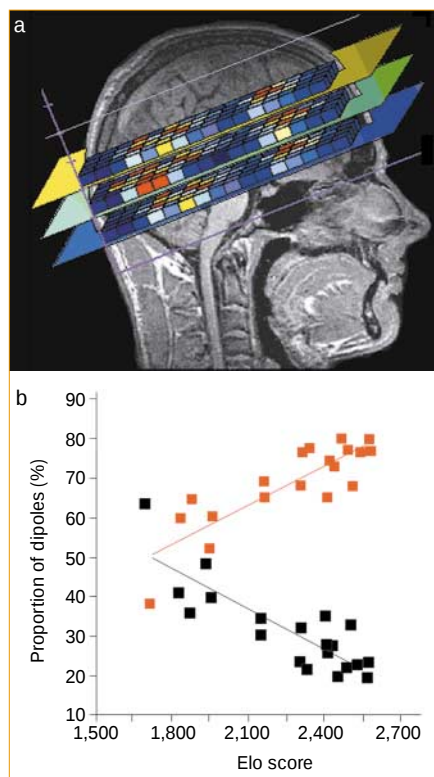


Figure 1 Focal γ -band activity in the brains of chess players. **a**, Determination of equivalent-current dipole density for inferior 'slices' through the hippocampus and medial temporal areas in an amateur player's brain. Increasing dipole density is indicated by a colour scale from dark to light blue, to yellow, to red. **b**, Relationship between chess-playing skill (Elo rating scale) and the relative share of dipoles located in medial temporal lobe structures (black) and in the frontal and parietal cortices (red). Amateur players show more focal γ -bursts in the medial temporal lobe than grandmasters, who show more activity in the frontal and parietal cortices.

Anabolism

Low mechanical signals strengthen long bones

Although the skeleton's adaptability to load-bearing has been recognized for over a century¹, the specific mechanical components responsible for strengthening it have not been identified. Here we show that after mechanically stimulating the hindlimbs of adult sheep on a daily basis for a year with 20-minute bursts of very-low-magnitude, high-frequency vibration, the density of the spongy (trabecular) bone in the proximal femur is significantly increased (by 34.2%) compared to controls. As the strain levels generated by this treatment are three orders of magnitude below those that damage bone tissue, this

anabolic, non-invasive stimulus may have potential for treating skeletal conditions such as osteoporosis.

A common perception of bone adaptation is that mechanical signals must be large to influence morphology². The peak signals that result from natural vigorous activity cause microdamage to bone material and require repair³. For example, peak strains of 2,000–3,000 microstrain are typically induced during locomotion⁴, stimulating osteoclasts and osteoblasts to remove and then replace damaged tissue⁵.

We have departed from this repair-mediated hypothesis by proposing that extremely small strains (for example, those that arise from muscle contraction during less vigorous but more frequent activities such as maintaining posture) are strong determinants of bone morphology⁶.

We examined the regulatory potential of extremely small (0.3g, where g is the Earth's gravitational field), high-frequency (30 Hz) mechanical accelerations by subjecting the hindlimbs of adult (6–8-year-old) female sheep (Warhill, intact ewes) to a ground-based vertical oscillation⁷ for 20 min per day for 5 days a week. When the animals were not being treated, they joined the controls to roam freely over a pasture area. We used strain gauges attached to the animals' tibia bone to calibrate the device: these showed that the peak-to-peak amplitude of the strain generated was about 5 microstrain, which is 0.1% of the strain magnitude that is known to cause yield-failure in bone⁸.

After 1 year of this mechanical-stimulation regime, the density of trabecular bone in the proximal femur, as quantified by computer tomography, was 34.2% greater in experimental sheep than in controls ($P < 0.01$; Table 1). This strong anabolic response was substantiated by undecalcified bone histology of the same region, which revealed a 32% increase in trabecular bone volume, a 45% increase in trabecular mesh number (Fig. 1) and a 36% reduction in mesh spacing, indicating an increase in the mean width of each trabecular element and the addition of new trabeculae.

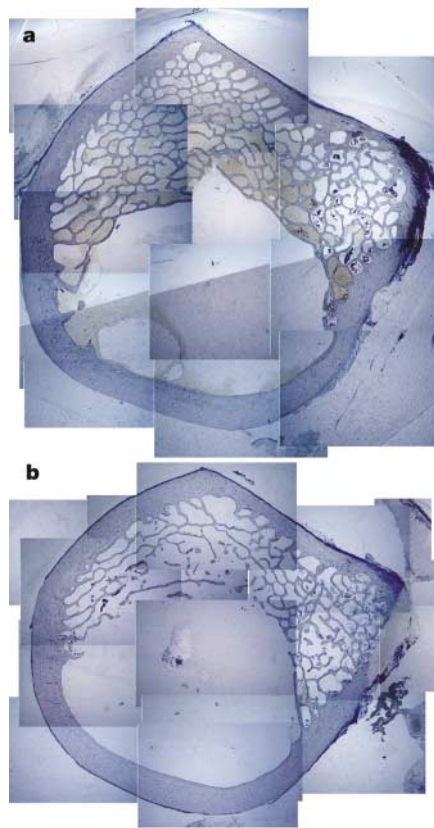


Figure 1 Montages of photomicrographs of the proximal sheep femur used for static histomorphometric evaluation after 1 year of exposure (20 min per day) to a 0.3g, 30-Hz mechanical stimulus. **a, b**, There is 32% more trabecular bone in the proximal femur of experimental animals (**a**) compared with age-matched controls (**b**) ($P < 0.04$).

Table 1 Proximal-femur parameters of control and stimulated sheep

	Control	Experimental	Difference	P
Animal mass (kg)	71.1 ± 7.1	70.3 ± 9.4	–1.1%	n.s.
Total density (gm cm ^{–3})	466 ± 60	496 ± 53	+6.5%	< 0.1
Trabecular density (gm cm ^{–3})	169 ± 37	227 ± 56	+34.2%	< 0.01
Bone volume/total volume (%)	15.2 ± 4.1	20.1 ± 4.8	+32%	< 0.04
Trabecular spacing (μm)	1,170 ± 124	756 ± 97	–36%	< 0.02
Trabecular number (trabeculae mm ^{–2})	0.82 ± 0.16	1.19 ± 0.18	+45%	< 0.01
Bone-formation rate (μm ³ mm ^{–2})	8.4 ± 12.7	17.9 ± 16.3	+113%	< 0.2
Mineralizing surface (%)	2.6 ± 0.16	6.34 ± 5.14	+144%	< 0.1

Animal mass and envelope-specific bone density (determined by quantitative computer tomography) of the proximal femur after 12 months of low-level mechanical stimulation. Also shown are indices of static and dynamic histomorphometry of the proximal femur. Although 'whole-bone' parameters of the proximal femur show only a limited tendency to be influenced by mechanical stimuli ($P = 0.1$), the increase in treated animals compared with controls is over 30% for trabecular bone alone ($P < 0.01$). One control was lost over the course of study for reasons not associated with the protocol. Nine animals were evaluated in the experimental group, with eight controls. All evaluations were made without knowledge of whether the animals were control or experimental.

We found that this low-level mechanical stimulation increased the rate of bone formation 2.1-fold ($P < 0.2$) and the mineralizing surface 2.4-fold ($P < 0.1$). This anabolic effect was highly specific to cancellous (porous) bone, as there was no significant histomorphometric change in any of the cortical bone parameters. We detected no difference in any bone index in the radius of either control or experimental animals (for example, mineral density was 0.6% less than controls; not statistically significant), indicating that the anabolic effect was specific to the region of the skeleton that was subjected to the mechanical signal.

Mechanical strain in the skeleton is a product of functional load-bearing — as seen, for example, in the mandible of the macaque⁹ and the tibia of the alligator¹⁰. In addition to the large-amplitude strains typically associated with functional activity, a strain signal, much less than 5 microstrain in amplitude, arises through muscular activity in the frequency band 10–50 Hz (ref. 11). Generation of this small-amplitude, high-frequency muscle 'vibration' persists through even such passive activities as standing.

Skeletal morphology may therefore be sculpted by omnipresent, low-level muscle activity as well as by the peak impacts inherent in load-bearing. In terms of clinical relevance, the strong bone-generating capacity of these small signals suggests that biomechanical intervention might help to strengthen bone in osteoporosis sufferers without the side-effects associated with pharmacological treatment. In addition to being non-invasive and inducing a therapeutic response from the bone tissue itself, low-intensity mechanical signals incorporate all aspects of a complex remodelling cycle¹² and ultimately stimulate formation of lamellar bone¹³ to improve bone quantity and quality.

Clinton Rubin*, **A. Simon Turner†**,
Steven Bain‡, **Craig Mallinckrodt†**,
Kenneth McLeod*

*Musculo-Skeletal Research Laboratory,
Department of Biomedical Engineering,
State University of New York, Stony Brook,

New York 11794-2580, USA

e-mail: clinton.rubin@sunysb.edu

†Department of Clinical Sciences, Colorado State
University, Fort Collins, Colorado 80523, USA

‡Skeletech Inc., Bothell, Washington 98021, USA

1. Wolff, J. *The Law of Bone Remodeling* (transl. Maquet, P. & Furlong, R.) (Springer, Berlin, 1986).
2. Frost, H. *Anat. Rec.* **26**, 403–413 (1990).
3. Burr, D., Martin, R., Schaffler, M. & Radin, E. *J. Biomech.* **18**, 189–200 (1985).
4. Rubin, C. & Lanyon, L. *J. Exp. Biol.* **101**, 187–211 (1982).
5. Schaffler, M., Radin, E. & Burr, D. *Bone* **11**, 321–326 (1990).
6. Huang, R., McLeod, K. & Rubin, C. *J. Gerontol.* **54**, 352–357 (1999).
7. Fritton, J., Rubin, C., Qin, Y. & McLeod, K. *Ann. Biomed. Eng.* **25**, 831–839 (1997).
8. Carter, D., Harris, W., Vasu, R. & Caler, W. *Am. Soc. Mech. Eng.* **45**, 81–95 (1981).
9. Hylander, W., Ravosa, M., Ross, C. & Johnson, K. *Am. J. Phys. Anthropol.* **107**, 257–271 (1998).
10. Blob, R. & Biewener, A. *J. Exp. Biol.* **202**, 1023–1046 (1999).
11. Fritton, S., McLeod, K. & Rubin, C. *J. Biomech.* **33**, 317–326 (2000).
12. Bain, S. & Rubin, C. *J. Bone Miner. Res.* **5**, 1069–1075 (1990).
13. Rubin, C., Gross, T., McLeod, K. & Bain, S. *J. Bone Miner. Res.* **10**, 488–495 (1995).

Vision

Realignment of cones after cataract removal

Through unique observations of an adult case of bilateral congenital cataract removal¹, we have found evidence that retinal photoreceptors will swiftly realign towards the brightest regions in the pupils of the eye. Cones may be phototropic, actively orientating themselves towards light like sunflowers in a field.

For 40 years, one of us (P.D.) lived with an unusual pupil configuration caused by congenital bilateral cataracts, first diagnosed at three years of age. The cataracts were managed by using a twice-weekly application of atropine, which dilated P.D.'s pupils to produce roughly annular (ring-shaped), clear regions around the dense nuclear cataracts. Because his optics were so poor (he was far-sighted, astigmatic and incapable of accommodation), P.D. squinted continuously as an adaptation, particularly for close work. His eyelids thus horizontally cropped each annulus at the top and bottom, leaving two clear entry points for light

Antarctic stratification and glacial CO₂

One way of accounting for lowered atmospheric carbon dioxide concentrations during Pleistocene glacial periods is by invoking the Antarctic stratification hypothesis, which links the reduction in CO₂ to greater stratification of ocean surface waters around Antarctica^{1,2}. As discussed by Sigman and Boyle³, this hypothesis assumes that increased stratification in the Antarctic zone (Fig. 1) was associated with reduced upwelling of deep waters around Antarctica, thereby allowing CO₂ outgassing to be suppressed by biological production while also allowing biological production to decline, which is consistent with Antarctic sediment records⁴. We point out here, however, that the response of ocean eddies to increased Antarctic stratification can be expected to increase, rather than reduce, the upwelling rate of deep waters around Antarctica. The stratification hypothesis may have difficulty in accommodating eddy feedbacks on upwelling within the constraints imposed by reconstructions of winds and Antarctic-zone productivity in glacial periods.

To suppress deepwater upwelling around Antarctica, it is necessary either to weaken the westerly wind stress sufficiently in the latitude band of Drake Passage to eliminate the northward flow (Ekman drift) of surface waters, or to counteract this northward drift through southward surface-eddy transport⁵. The first possibility does not seem feasible for glacial periods: the Ekman drift across 55° S is ~25 Sv (1 Sv = 10⁶ m³ s⁻¹; ref. 6) today, and was almost certainly greater during glacial periods because of stronger winds, as expected from the increased Equator-to-Pole temperature gradient and supported by the increased transport of sea salt to the Antarctic continent⁷. This conclusion probably holds even if the latitude of maximum westerly winds shifted slightly towards the Equator, as Sigman and Boyle speculate³.

The second possibility is also problematic if the glacial Antarctic surface ocean was more stratified than today: eddy transports are thought to scale in proportion to the degree of baroclinicity, which is measured by the slope and spacing of the isopycnal surfaces. High baroclinicity favours vigorous eddy transport, whereas low baroclinicity (that is, stratified conditions) suppresses eddy transport⁸. Today, the Antarctic deep upwelling and associated net northward surface flow is estimated to be ~10 Sv, on the basis of geochemical constraints on the partitioning of deep

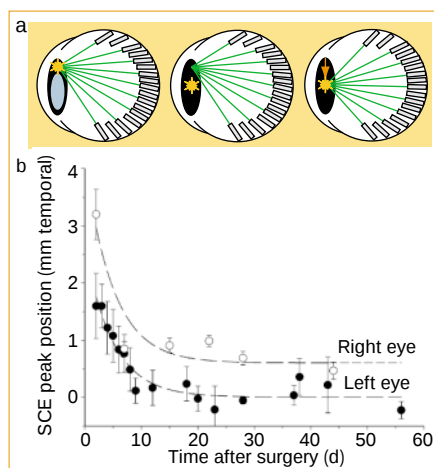


Figure 1 Photoreceptor alignment in P.D.'s eyes before and after cataract removal. **a**, Schematic horizontal cross-sections through P.D.'s eyes. Left, before surgery: receptors were aligned (green lines) with the brightest pupil location (orange 'sun') in the temporal margins; middle, immediately after surgery: the brightest location shifted to the pupil centres, but the receptors retained their skewed alignment; right, after 10 days: receptors realigned with the bright pupil centres (arrow depicts shifts measured in **b**). **b**, Photoreceptor alignment in the pupil plane after cataract removal, as inferred from Stiles–Crawford effect (SCE I) peaks, for left and right eyes.

in each eye. Because of the asymmetrical locations of the cataracts within the dilated pupils, the clear regions furthest from the nose (temporal) were wider and larger than those closest to the nose (nasal; Fig. 1a, left).

In adults with normal pupils, receptors are aligned towards the pupil centres, where light is brightest^{2,3}. They are aligned there from birth and remain so for life². As part of a wider characterization of his visual system¹, we noted that P.D.'s photoreceptors were aligned with the larger, temporal clear regions in each eye. We made this discovery when P.D. spontaneously reported that point sources of monochromatic red laser light presented in the temporal clear regions appeared distinctly “brighter and whiter”. Differences in the brightness and colour of light entering the pupil from different points are termed Stiles–Crawford effects of the first and second kind (SCE I and SCE II), respectively^{4,5}, and reflect the alignment of cone photoreceptors^{2,6}. These effects result from the waveguide properties of photoreceptors, which endow them with directional selectivity for incident light².

P.D.'s decision to have his cataracts removed gave us a unique opportunity to test the proposal that cones are actively phototropic, dynamically maintaining alignment with the brightest region of the pupil^{7–10}. This was because surgery would shift the brightest regions of his pupils from the temporal margins to the centres (Fig. 1a, middle). The right cataract was removed six weeks after the left one.

We measured SCE I functions across the central horizontal pupil meridians every

day after surgery¹. Figure 1b shows that the peaks of the SCE I functions migrated nasally to the pupil centres over a 10-day period. The left-eye peak moved 1.6 mm, equivalent to 4° of receptor realignment. The right-eye peak moved further (2.6 mm, 6.5°) but did not entirely re-centre, stabilizing instead 0.6 mm to the temporal side. On this evidence, photoreceptors seem to be phototropic (Fig. 1a, right).

P.D. had two separate entry points for light in each eye before surgery — could there have been a concomitant dual photoreceptor alignment, with a minority of receptors aligned with the smaller, nasal clear regions? We failed to find support for this idea, as adapting lights presented to the two clear regions had no differential effect on sensitivity¹¹, suggesting that the receptors all had a common alignment; moreover, the spread of SCE I function (an index of receptor disarray⁶) was normal and did not decrease after surgery.

The rate of photoreceptor realignment was proportional to the distance remaining to the pupil centres. The dashed curves in Fig. 1b are exponentials that fit best with 5-day time constants for both eyes. Thus, the biophysical processes that underlie phototropism may be under the control of a simple feedback signal. In plants, phototropism is mediated by differential growth¹². In eyes, it may be that differential longitudinal growth mediated by actin filaments or microtubules makes each photoreceptor phototropic^{8,13}. If this is correct, then our analogy with sunflowers is not trivial.

Harvey S. Smallman*†, Donald I. A. MacLeod*, Peter Doyle‡

*Department of Psychology, University of California at San Diego, La Jolla, California 92093, USA

†Pacific Science and Engineering Group, San Diego, California 92122, USA

e-mail: smallman@pacific-science.com

‡Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755, USA

1. Fine, I., Smallman, H. S., Doyle, P. & MacLeod, D. I. A. *Vision Res.* (in the press).
2. Enoch, J. M. & Lakshminarayanan, V. J. in *Visual Optics and Instrumentation* (ed. Charman, W. N.) 280–309 (CRC, Boca Raton, Florida, 1991).
3. Laties, A. M., Lieberman, P. & Campbell, C. *Nature* **218**, 172–173 (1968).
4. Stiles, W. S. & Crawford, B. H. *Proc. R. Soc. Lond. B* **112**, 428–450 (1933).
5. Enoch, J. M. & Stiles, W. S. *Optica Acta* **8**, 329–358 (1961).
6. Applegate, R. A. & Lakshminarayanan, V. J. *Opt. Soc. Am. A* **10**, 1611–1623 (1993).
7. Bonds, A. B. & MacLeod, D. I. A. *Invest. Ophthalmol. Vis. Sci.* **17**, 754–761 (1978).
8. Applegate, R. A. & Bonds, A. B. *Invest. Ophthalmol. Vis. Sci.* **21**, 869–872 (1981).
9. Enoch, J. M. & Birch, D. G. *Phil. Trans. R. Soc. Lond. B* **291**, 323–351 (1981).
10. Kono, M. et al. *Vision Res.* **41**, 103–118 (2001).
11. MacLeod, D. I. A. *Vision Res.* **14**, 369–374 (1974).
12. Dennison, D. S. in *Encyclopedia of Plant Physiology, New Series* (eds Haupt, W. & Feinlib, M. E.) 506–566 (Springer, Berlin, 1979).
13. Laties, A. M. & Burnside, B. in *Motility and Cell Function* (eds Pepe, F., Nachmias, V. & Sanger, J. W.) 285–298 (Academic, New York, 1979).

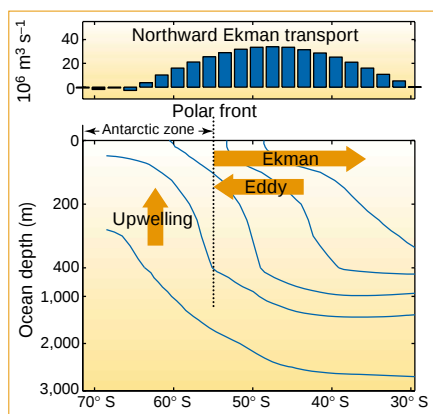


Figure 1 Northward Ekman transport of surface waters and contours of modern potential density averaged for each ocean zone. The approximate mean latitude of the Polar Front (northern boundary of Antarctic zone) and direction of Ekman and eddy transports that control the rate of deepwater upwelling are shown (lower panel). The relevant eddy transport involves a net southward flow, as opposed to north–south mixing, and ultimately results from the tendency of lighter water to spread over denser water.

upwelling between low and high latitudes^{9,10}; this is consistent with scaling arguments⁵ and global inversions¹¹. Modern southward surface-eddy transport must therefore be of similar magnitude (~15 Sv). In a more stratified glacial ocean, the eddy transport in the surface layer would weaken and upwelling would consequently increase, an opposite change to that suggested by Sigman and Boyle.

The stratification hypothesis could survive in the face of eddy feedbacks if glacial winds were markedly weakened in the relevant latitude band, or if the biological productivity of the Antarctic zone increased sufficiently to prevent CO₂ outgassing despite greater upwelling. However, neither change seems probable given the available evidence. The primary cause of lowered glacial CO₂ may not have been enhanced year-round stratification, but enhanced Antarctic sea-ice coverage¹², which could have suppressed CO₂ outgassing even if glacial Antarctic surface waters were generally less stratified than today, with upwelling into a buoyant surface layer confined to a brief summer season¹³.

Ralph F. Keeling*, Martin Visbeck†

*Scripps Institution of Oceanography, La Jolla, California 92093-0244, USA
e-mail: rkeeling@ucsd.edu

†Lamont-Doherty Earth Observatory, Palisades, New York 10964-8000, USA

1. Francois, R. F. *et al.* *Nature* **389**, 929–935 (1997).
2. Toggweiler, J. R. *Paleoceanography* **14**, 571–588 (1999).
3. Sigman, D. M. & Boyle, E. A. *Nature* **407**, 859–869 (2000).
4. Mortlock, R. A. *et al.* *Nature* **351**, 220–223 (1991).
5. Gnanadesikan, A. *Science* **283**, 2077–2079 (1999).
6. Trenberth, K. E., Large, W. G. & Olson, J. G. *J. Phys. Oceanogr.* **20**, 1742–1760 (1990).
7. Petit, J., Briat, M. & Royer, A. *Nature* **293**, 391–394 (1981).
8. Visbeck, M., Marshall, J., Haine, T. & Spall, M. *J. Phys. Oceanogr.* **27**, 381–402 (1997).
9. Toggweiler, J. R. & Samuels, B. in *The Global Carbon Cycle* (ed.

- Heimann, M.) 303–331 (Springer, Berlin, 1993).
10. Gnanadesikan, A. & Toggweiler, J. R. *Geophys. Res. Lett.* **26**, 1865–1868 (1999).
11. Ganachaud, A. & Wunsch, C. *Nature* **408**, 453–456 (2000).
12. Stephens, B. B. & Keeling, R. F. *Nature* **404**, 171–174 (2000).
13. Keeling, R. F. & Stephens, B. B. *Paleoceanography* **16**, 330–334 (2001).

Sigman and Boyle reply — Palaeoceanographic evidence indicates that there was more complete nutrient consumption in Antarctic surface waters during the last ice age^{1,2}, but lower biological production¹. These results suggest that the Antarctic was stratified during glacial times, reducing the transport of sequestered nutrients and CO₂ into the Antarctic surface. By sequestering CO₂ in the ocean interior, this change could explain the observation of lower levels of atmospheric CO₂ during the ice age³. Geological data offer two possible causes for this stratification. First, the Southern Hemisphere westerly winds apparently shifted northwards during glacial times⁴, which would have reduced Ekman-driven upwelling in the Antarctic⁵ (a ‘wind-shift’ mechanism). Second, the Antarctic sea-ice cycle intensified during glacial times⁶, which may have allowed a low-salinity lid to accumulate in the open Antarctic, thus reducing vertical mixing and open-ocean overturning (a ‘sea-ice’ mechanism).

Keeling and Visbeck criticize these mechanisms for Antarctic stratification on theoretical grounds and highlight an alternative hypothesis for lowering glacial CO₂ — prevention of CO₂ release from the Antarctic by covering the ocean with sea ice, thereby blocking ocean–atmosphere CO₂ exchange⁷. Although we cannot be completely confident about the specific mechanisms for stratification outlined above, we believe that Antarctic stratification is a more plausible hypothesis for lower glacial CO₂ than gas-exchange limitation, and it is also more directly supported by palaeoceanographic data³.

With regard to the wind-shift mechanism, Keeling and Visbeck argue that a reduction in winds over the Antarctic was unlikely because of an increase in the Equator-to-Pole temperature gradient during glacial times. However, the modern meridional variation in wind strength across the Southern Ocean is large enough for the observed northward migration in westerly winds during the ice age to have overcome the effects of a global average increase in winds, yielding less wind-driven upwelling in the glacial Antarctic. Winds depend on regional (not global) temperature gradients, and the temperature gradient across the Antarctic may well have been smaller during glacial times, potentially explaining the greater northward persistence of sea ice. But it must be admitted that the wind-shift mechanism is

complex and has inherent thresholds⁸, so this mechanism may have difficulty in accounting for the timing of CO₂ change and its robust, linear relationship with Southern Hemisphere temperature^{4,9}.

Keeling and Visbeck criticize the sea-ice mechanism for stratification on the grounds that it would have been countered by an increase in upwelling because of a response in the southward flux of eddies to a change in ocean-density structure. However, the eddy response is a negative feedback which, at most, would set boundaries on the stratification caused by the sea-ice mechanism. Moreover, the full slope of the density surfaces at the polar front might have changed very little if only the shallowest 30–50 m of the Antarctic surface stratified, in which case there would have been no eddy response. In our opinion, a greater problem with the sea-ice mechanism involves higher-latitude conditions: stratification of the open Antarctic as a result of an enhanced sea-ice cycle might occur at the expense of the coastal Antarctic, making this region more saline and thus more active in ocean ventilation, with an accompanying release of CO₂ into the atmosphere.

Prevention of ocean–atmosphere CO₂ exchange in the Antarctic by sea-ice cover⁷ is unlikely to be the sole mechanism for reducing CO₂ levels during ice ages, because it would require almost complete and continuous ice coverage of the region. For this reason, Keeling and Visbeck refer to the previously described¹⁰ hybrid hypothesis that invokes intense surface stratification and nutrient consumption during the summer, followed by prevention of gas exchange by ice cover during winter. Although promising, this mechanism faces a discrepancy with the evidence of lower productivity in the glacial Antarctic. Without permanent stratification, greater nutrient consumption, even for a brief summer period, would have required a larger annual export of organic matter from the Antarctic surface.

Daniel M. Sigman*, Edward A. Boyle†

*Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA

e-mail: sigman@princeton.edu

†Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

1. Francois, R. F. *et al.* *Nature* **389**, 929–935 (1997).
2. Sigman, D. M., Altabet, M. A., Francois, R., McCorkle, D. C. & Gaillard, J.-F. *Paleoceanography* **14**, 118–134 (1999).
3. Sigman, D. M. & Boyle, E. A. *Nature* **407**, 859–869 (2000).
4. McCulloch, R. D. *et al.* *J. Quat. Sci.* **15**, 409–417 (2000).
5. Toggweiler, J. R., Carson, S. & Bjornsson, H. *EOS Trans. Am. Geophys. Un.* **80**, OS286 (1999).
6. Crosta, X., Pichon, J.-J. & Burckle, L. H. *Paleoceanography* **13**, 284–297 (1998).
7. Stephens, B. B. & Keeling, R. F. *Nature* **404**, 171–174 (2000).
8. Toggweiler, J. R. & Samuels, B. *Deep-Sea Res.* **1**, 42, 477–500 (1998).
9. Monnin, E. *et al.* *Science* **291**, 112–114 (2001).
10. Moore, J. K., Abbott, M. R., Richman, J. G. & Nelson, D. M. *Glob. Biogeochem. Cycles* **14**, 455–475 (2000).

Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair

John R. Walker^{*†}, Richard A. Corpina^{*†} & Jonathan Goldberg^{*‡}

^{*} Cellular Biochemistry and Biophysics Program and [‡] Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

[†] These authors contributed equally to this work.

The Ku heterodimer (Ku70 and Ku80 subunits) contributes to genomic integrity through its ability to bind DNA double-strand breaks and facilitate repair by the non-homologous end-joining pathway. The crystal structure of the human Ku heterodimer was determined both alone and bound to a 55-nucleotide DNA element at 2.7 and 2.5 Å resolution, respectively. Ku70 and Ku80 share a common topology and form a dyad-symmetrical molecule with a preformed ring that encircles duplex DNA. The binding site can cradle two full turns of DNA while encircling only the central 3–4 base pairs (bp). Ku makes no contacts with DNA bases and few with the sugar-phosphate backbone, but it fits sterically to major and minor groove contours so as to position the DNA helix in a defined path through the protein ring. These features seem well designed to structurally support broken DNA ends and to bring the DNA helix into phase across the junction during end processing and ligation.

Damage to DNA in the form of double-strand breaks (DSBs) compromises the integrity of cells, and defects in DSB repair processes can lead to chromosomal translocations, loss of growth control and cancer¹. DSBs are repaired by either homologous recombination or non-homologous end joining (NHEJ). In mammalian cells, NHEJ is a principal pathway for the repair of DSBs that result from DNA-damaging agents such as ionizing radiation, and is required for repair of the programmed DSBs that arise during V(D)J recombination in lymphocytes².

Genetic and biochemical experiments identified a core set of proteins, conserved from yeast to mammals, which mediate repair by the NHEJ pathway. These are the 70K and 80K subunits (relative molecular masses (M_r) of 70,000 and 80,000, respectively) of the Ku heterodimer (Ku70, Ku80), and the complex of XRCC4 and DNA ligase IV (refs 3, 4)—vertebrate cells also require the catalytic subunit of DNA-dependent protein kinase (DNA-PK_{cs})^{5,6}. Although NHEJ joins broken ends directly without a template, it can be a relatively accurate process that tends to preserve the sequence at the junction. The Ku heterodimer is required for accurate NHEJ, as demonstrated by the high frequency of imprecise end joining in cells and extracts that lack Ku^{7–10}. The importance of Ku for DSB repair and genetic stability is underscored by the extreme radiation sensitivity and specific V(D)J recombination defects of Ku-deficient cells^{11–14}, and by the high levels of chromosomal aberrations observed in cells of Ku70^{−/−} and Ku80^{−/−} mice^{15,16}.

Studies of the NHEJ reaction suggest that DNA termini are held together precisely in the repair complex, allowing polymerases to fill in gaps and nucleases to trim excess ends as necessary, before ligation—as few as 1–2 complementary base pairs, if present, can position the junction^{17,18}. The capacity of vertebrate cells to join DNA ends in this way exceeds that of individual ligase and polymerase activities, and it has been inferred that accurate repair by NHEJ is potentiated by alignment factors that structurally support and align DNA ends to maintain the correct setting of thermodynamically weak base pairing and stacking interactions^{9,18}. Biochemical studies suggest that the Ku heterodimer recognizes DSBs and acts as an alignment factor to promote end joining^{9,19–21}. Ku is an abundant nuclear protein that binds with high affinity to duplex DNA ends (and can also translocate internally on a DNA fragment),

independent of the end sequence or precise structure, but binds with low affinity to circularized DNA²². Once bound to DNA, Ku can recruit DNA-PK_{cs}, which seems to have a regulatory role in the end-joining process involving its protein kinase activity^{10,23}. Current models suggest that on DSB formation, each broken terminus binds a Ku heterodimer and the two heterodimers then associate to form a bridging complex that recruits additional NHEJ proteins^{19,20}. The ability of DNA-bound Ku to self-associate has been demonstrated both biochemically²⁰ and by atomic force microscopy imaging¹⁹. More directly, Ku stimulates ligation by mammalian DNA ligases *in vitro*, which is consistent with a role for Ku in bridging two DNA ends yet maintaining accessibility of the ends to the ligase²⁰.

The topological principles of DSB recognition and alignment-based end joining are currently unknown; however, the DNA-binding function of Ku is clearly fundamental to the process. By designing a DNA substrate with a single accessible end it was possible to isolate and crystallize an oriented complex comprising the human Ku heterodimer bound to DNA. Its structure is described here at 2.5 Å resolution; for comparison, the crystal structure of Ku in the absence of DNA is presented at 2.7 Å resolution.

Structure determination

All structural studies used the heterodimer comprising a full-length Ku70 and a truncated Ku80, which lacks a carboxy-terminal 19K domain known to function in DNA-PK recruitment, but which is dispensable for DNA end binding^{24,25}. Crystallization experiments in the absence of DNA yielded orthorhombic crystals that diffracted X-rays to at least 2.5 Å resolution (space group $P2_12_12_1$; $a = 80.2$ Å, $b = 86.2$ Å, $c = 203.3$ Å). The structure was determined by the multiple-wavelength anomalous diffraction (MAD) method using selenium as the anomalous scatterer (Supplementary Information Table 1).

As Ku has no sequence specificity when binding to DNA ends²², the main difficulty in obtaining crystals of a Ku–DNA complex was to prevent Ku from adopting multiple binding modes on a DNA substrate. Conversely, Ku is expected to load onto a DNA end with a defined polarity to properly organize the repair process. *In vitro* experiments confirmed that Ku indeed adopts a strongly preferred

orientation on a DNA probe bearing a single free end²⁶. This finding was utilized to force Ku towards a single DNA-binding mode by blocking one end of a 14-bp DNA duplex with a three-way junction comprising a stem loop plus a short G-rich stem (Fig. 1). The 150K complex of Ku bound to the 55-nucleotide DNA formed orthorhombic crystals (space group $P2_12_12_1$; $a = 83.9$ Å, $b = 126.3$ Å, $c = 126.4$ Å) that diffracted X-rays beyond 2.3 Å resolution. The structure was determined by the molecular replacement method and refined against data to 2.5 Å resolution (Supplementary Information Table 1). The nucleotide sequence and positions were verified by difference Fourier analysis of data from crystals containing oligonucleotides substituted at several sites with either iodouracil/cytosine or bromo-U/C.

General features of the Ku heterodimer

The fact that Ku forms a ring to encircle duplex DNA explains why Ku requires DNA ends for high-affinity binding (Figs 1 and 2). Unlike the uniform, symmetrical protein rings observed in replication processivity factors²⁷, the Ku ring is designed with an expansive base that cradles DNA, and a very narrow bridge (strand β J on each subunit) that acts as a barrier to promiscuous binding to unbroken DNA (Fig. 2). The folds of Ku70 and Ku80 are closely related (Fig. 3b) and the two form a quasi-symmetrical molecule, indicating that they diverged from an ancestral homodimer. This correspondence was identified previously on the basis of sequence homology^{22,28} (see Fig. 3 for a summary of combined sequence and structural information). The low level of sequence identity (~15%) among residues that contribute to the dimer interface should ensure heterodimer formation and preclude Ku70–70 or Ku80–80 homodimer formation. Ku70 and Ku80 share a three-domain topology comprising an amino-terminal α/β domain, a central β -barrel domain and a helical C-terminal arm (Figs 2 and 3). In addition, the subunits have divergent C-terminal regions: the 19K DNA-PK recruitment element of Ku80 (refs 25, 28), which is absent from the crystal structure, and a 5K region of Ku70, present in the crystals, identified previously as a SAP domain²⁹. The heterodimer has overall dimensions of $\sim 120 \times 70 \times 60$ Å. The α/β domains are distant from the protein dyad axis and define the long molecular axis, whereas the β -barrels are central and form the 70 Å-long base or cradle of the DNA-binding groove. Finally, the short axis is measured along the molecular dyad from the bottom of the β -barrel to the bridge across the DNA (Figs 1 and 2).

The N-terminal α/β domains (Ku70, residues 34–250; Ku80, residues 6–238; 13% sequence identity) lie at the periphery of the heterodimer and make only small contributions to the dimer interface. The domain comprises a six-stranded β -sheet of the

Rossmann fold type, but with one of the strands (β C) antiparallel (Figs 2c and 3). The amino edge of the β -sheet is proximal to the DNA-binding groove. The carboxy edge of the sheet is distal and has no role in DNA binding, but this is frequently a functionally important site in α/β proteins, and is potentially a site on each subunit for binding to other repair factors.

The β -barrel domains of Ku70 and Ku80 (11% sequence identity among 116 residues) are centred around the protein dyad and form contacts across the dimer interface (Figs 1 and 2). The two barrels have very similar conformations (average deviation of 0.95 Å after overlapping 97 C α atoms). The domain is constructed from a core of seven β -strands, arranged in an antiparallel fashion with the exception of the FN parallel pair (Fig. 2c). The use of seven β -strands yields a quite symmetrical, round barrel, unlike the flattened barrel of the immunoglobulin fold. The β -barrel can be separated into two smaller elements: the four C-terminal β -strands, OPQR, form a greek key motif that lies closest to the molecular dyad; the three N-terminal β -strands, FGN, include the parallel FN pair and a marked ~ 70 -residue insertion between strands β G and β N, which forms the ring around DNA (Figs 2 and 4). Together, the Ku70 and Ku80 β -barrels form the cradle of the DNA-binding site by pointing one end of each barrel towards the DNA (Fig. 2b). This is similar to the Rel-homology regions of NF- κ B, where loops at the ends of immunoglobulin-like β -barrels make sequence-specific contacts with the DNA major groove^{30,31}. However, the Ku β -barrel introduces just one loop (between strands β P and β Q) into the DNA minor groove (Fig. 5b), and the domain provides an otherwise relatively flat surface for DNA binding.

Topology of the Ku ring

Figure 2 shows the route of the polypeptide chain around DNA (Ku70, residues 277–341; Ku80, residues 267–336), and illustrates how this unusual structural element arises from a connection between two strands (GN) of the β -barrel. Ku70 and Ku80 share this feature, so that in effect Ku forms two rings around DNA. This arrangement of a polypeptide chain fully encircling DNA is unlike the replication processivity factors, where an oligomeric ring is constructed from globular protomers that each contribute just a portion of the ring²⁷. Evidently, Ku will become trapped on DNA after a repair event, which raises an important question concerning the mechanism of unloading.

Analysis of the fold and quaternary structure shows that unloading of a trapped Ku heterodimer from DNA might be achieved if the subunits could dissociate and thereby compromise the stability of the interface between α/β and β -barrel domains within a single subunit, involving some local unfolding of the β -barrel. However,

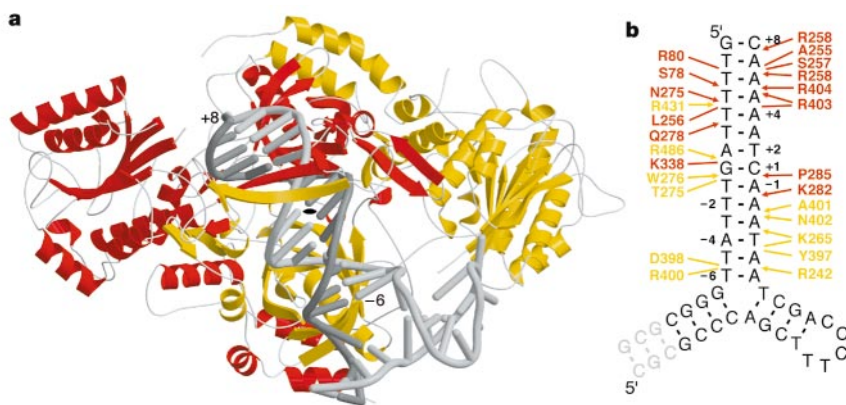


Figure 1 Overview of the crystal structure. **a**, Ribbon diagram of Ku bound to DNA, viewed down the molecular dyad axis (black symbol). Ku70 is coloured red and Ku80 orange. The 34-residue oligonucleotide is light grey; the 21-residue oligonucleotide dark grey. The terminal base pairs of the central duplex are numbered +8 (broken DNA end) and -6.

b, Secondary structure of the DNA used for crystallization. Base pairs in the central duplex are numbered from -6 to +8. Arrows (colour-coded as in **a**) indicate contacts between the protein and phosphates on DNA; lines indicate contacts to sugars. Three base pairs (grey) are disordered in the crystal.

Ku70 and Ku80 associate tightly in solution, and the structure reveals an interdigitated and spatially extensive interface ($\sim 9,000 \text{ \AA}^2$ of molecular surface area buried per subunit; see Fig. 3a). In particular, following the β -barrel, the polypeptide chain of each subunit stretches $\sim 40 \text{ \AA}$ across the DNA-binding groove, and forms the helical C-terminal arm (Ku70, residues 440–534; Ku80, residues 427–539; 17% sequence identity), which embraces the β -barrel of the opposite subunit (Figs 2c, 3c and 6).

A preformed channel for DNA

By forming a ring around DNA, Ku can achieve high-affinity binding (dissociation constant (K_d) values between 1.5 to $4.0 \times 10^{-10} \text{ M}$; see ref. 22) without recourse to sequence-specific bonding interactions. Two features of the protein enhance this effect: there is a polarization of positive electrostatic charge focused on the inner surface of the ring and along the DNA-binding cradle (Fig. 4b), and the ring is preformed for DNA binding.

Comparison of the DNA-free and -bound crystal structures shows that the ring maintains its conformation in the absence of DNA (Figs 2a and 6). The residues of the central polypeptide rings can be superimposed to a close fit (average deviation of $0.58 \text{ \AA}$ for

123 $\text{C}\alpha$ atoms), and this correspondence extends throughout the heterodimer (average deviation of $0.79 \text{ \AA}$ for 999 $\text{C}\alpha$ atoms). The stability of the protein ring is conferred by pillars positioned on either side of the DNA, which lie slightly inside the major groove (Fig. 2). The pillars are formed by the Ku70 and Ku80 polypeptide rings folding into a set of short β -hairpins (β -strand pairs HI, JK and LM), which then stack in an orthogonal, interleaved arrangement, three hairpins per pillar. To buttress the pillars, helix $\alpha 15$ of the C-terminal arm is wedged between the pillar and the α/β domain (Figs 2c and 6).

Figure 1b shows the DNA used for crystallization, including 3 bp of the G-rich stem that are disordered in the crystals. Several unpaired bases in the stem-loop make contacts with protein residues, either within the intermolecular complex or as crystal contacts. The stem-loop and G-rich portions of DNA are omitted from all other figures, which focus on the central 14-bp duplex bound by Ku. The DNA is positioned with the major groove lying centrally over the β -barrel cradle, such that the protein dyad axis coincides closely with a DNA dyad centred between base pairs -1 and $+1$ (Figs 1 and 2; see Fig. 1b for numbering scheme). The 14-bp region has a B conformation and in Fig. 5a additional B DNA has

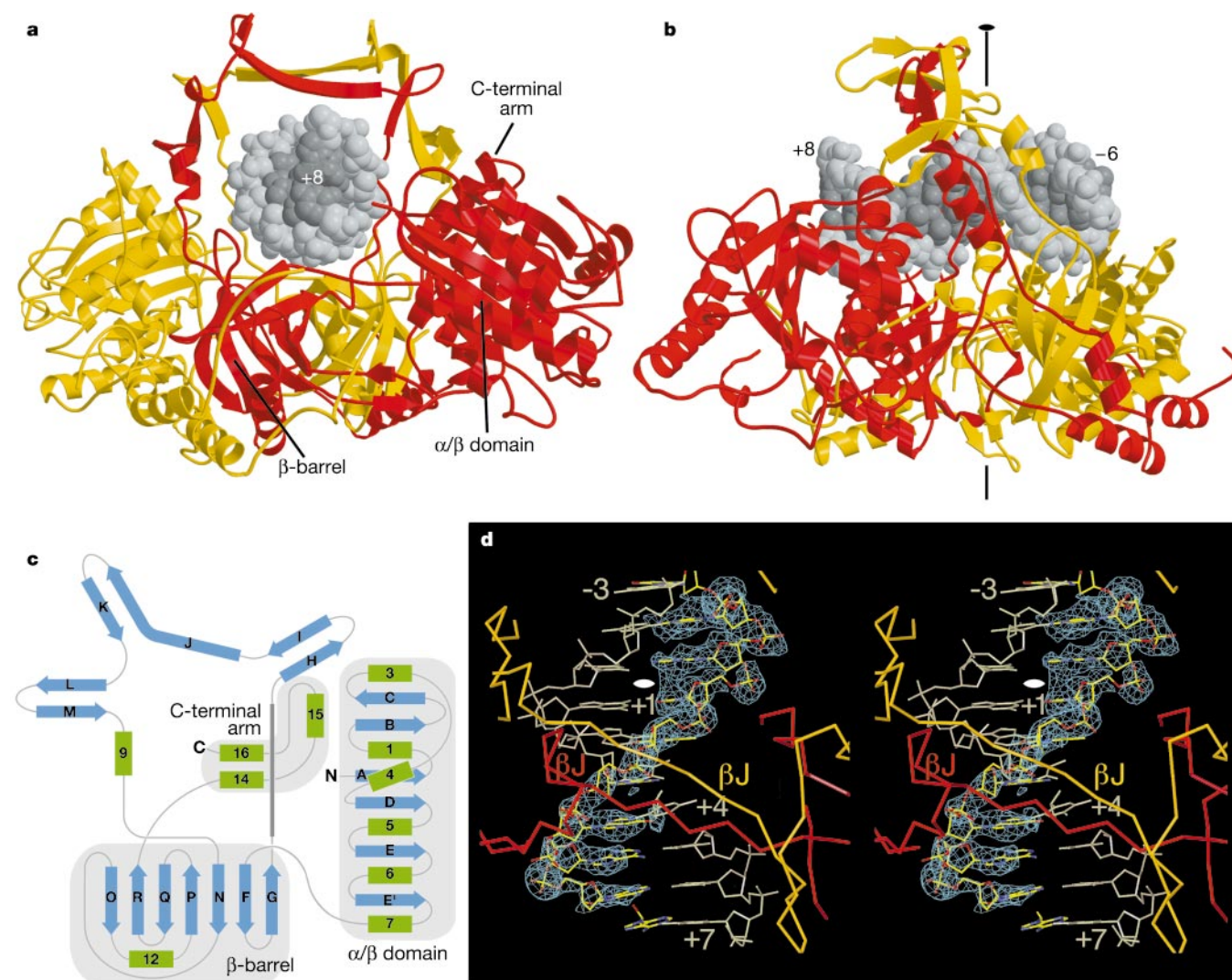


Figure 2 Structure of the Ku–DNA complex. **a**, View down the DNA helix. Ku70 is coloured red and Ku80 orange. Only the 14-bp duplex portion of DNA is shown; the sugar-phosphate backbone is coloured light grey and the bases dark grey. **b**, Side view. Black line indicates the molecular dyad axis. Terminal base pairs are numbered +8 (broken DNA end) and –6. **c**, Abbreviated topology diagram showing the fold of Ku70 and Ku80. Green

rods indicate α -helices; blue arrows indicate β -strands. **d**, Stereo difference electron density map viewed down the molecular dyad axis (white symbol). Blue contour lines show electron density at the 2.4σ -level in a $|F_o| - |F_c|$ ($2.5 \text{ \AA}$ resolution) map calculated after simulated-annealing refinement of a model lacking nucleotides from levels -2 to $+6$ of the 34-residue oligonucleotide.

been modelled to show the extent of the binding site. Two turns of DNA, roughly 20 bp, can fit along the 70 Å cradle. This is consistent with a repeat of 25–30 bp when several Ku heterodimers translocate internally and coat DNA³², and with neutron-scattering studies showing a 24-bp DNA element lying almost entirely within the Ku protein envelope³³.

Ku heterodimer binds to DNA ends in a preferred orientation, and photo-crosslinking studies indicate that Ku70 is located proximal, and Ku80 distal, to the free end²⁶. The crystal structure shows that the entrance to the DNA-binding site on one side is constructed almost entirely from Ku70 residues, and on the other side by Ku80 residues (Figs 2b and 4a), and that the free DNA end lies on the Ku70 side (defined here as the + end). The asymmetry in DNA binding may arise from one or a combination of several asymmetric structural features of the heterodimer. (These features are not due to crystal packing forces or to the DNA three-way junction, as they

are observed in both crystal structures.) First, although the bridge and pillar regions have a similar topology in the two subunits, this structure as a whole is shifted away from the dyad axis and towards the Ku70 end of the DNA-binding site (Fig. 2b). Second, the α/β domains deviate from proper symmetry, being related by a 168° rotation. The effect is to position the Ku80 α/β domain further from DNA. The Ku70 α/β domain is ~7 Å closer so that the N termini of strands β A, β B and β D abut the DNA in the vicinity of base pairs +5 to +8 (Fig. 5a). The N terminus of Ku70 borders the DNA-binding groove, and the 33 N-terminal residues are disordered in both crystal forms. This is a highly acidic peptide (12 out of 33 residues are aspartate or glutamate), and it may provide a partial steric and electrostatic block to DNA access on the Ku70 side. Finally, in addition to conferring a preferred orientation on DNA binding, these asymmetric features may also impose an energetic barrier to inward movement of Ku on DNA, and thereby help to retain Ku at

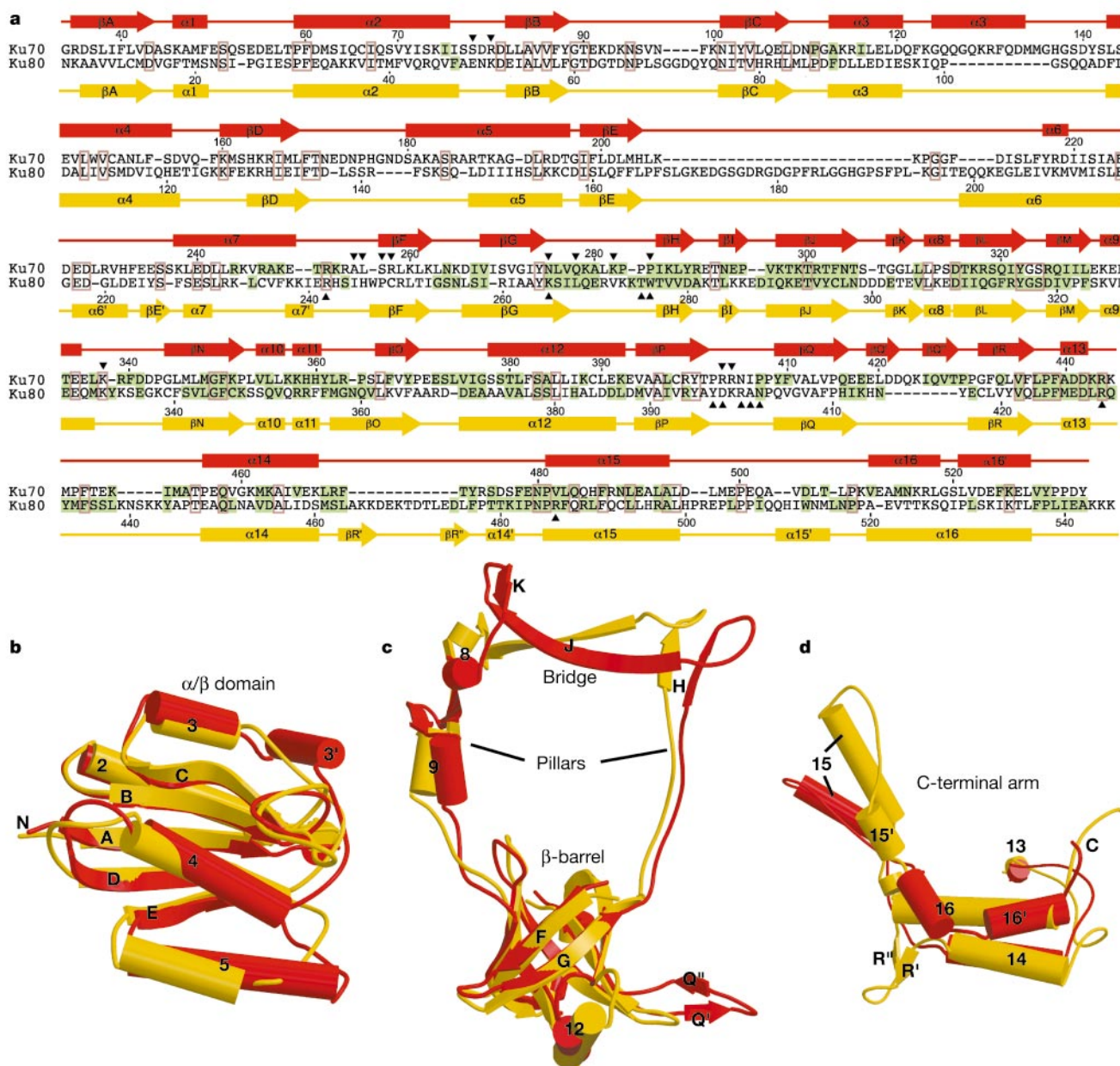


Figure 3 Relatedness of human Ku70 and Ku80. **a**, Structure-based sequence alignment (14% identity). Secondary structure elements are indicated in red for Ku70 and orange for Ku80. Cylinders indicate α -helices; arrows indicate β -strands. Residues making inter-subunit interactions are highlighted in green, those making DNA contacts are indicated

with a black arrowhead, and invariant residues are outlined in brown. **b**, Overlay of the α/β (Rossmann fold) domains of Ku70 (red) and Ku80 (orange). The polypeptide chains diverge beyond strand β E, so those regions are omitted for clarity. **c**, Overlay of the β -barrels and polypeptide rings. **d**, C-terminal arm regions.

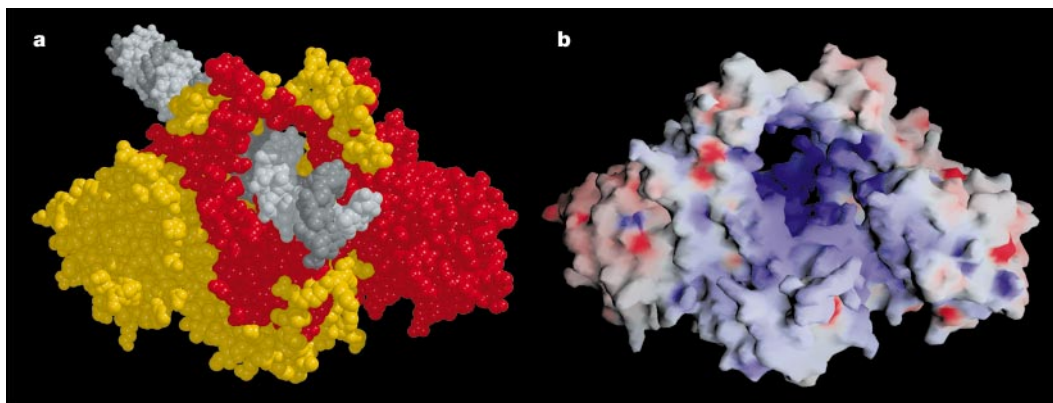


Figure 4 Surface depictions of Ku. **a**, Space-filling model showing Ku bound to DNA. The model was prepared by fitting a 32-bp B DNA to the crystallographically observed duplex. DNA extends towards the viewer to the +11 level. Ku70 is coloured red and Ku80 orange.

DNA is shown with one light grey and one dark grey strand. **b**, Molecular surface representation of Ku is coloured according to electrostatic potential, calculated using the program GRASP⁴⁴. Negative potential is coloured red and positive potential blue.

the DNA end. In this view, the preference of Ku for binding DNA ends arises from a slow dissociation from the end, consistent with biochemical experiments demonstrating inward translocation of Ku as the rate-limiting step in binding to internal DNA sequences³².

A topological problem that arises in the NHEJ reaction is how the two DNA ends are brought together in an alignment complex yet remain accessible to enzymes that carry out fill-in synthesis, nucleolytic processing and ligation². Recent biochemical studies indicate that this ability is an inherent property of the Ku hetero-

dimer itself, on the basis of its capacity to promote end-to-end interactions and stimulate ligation²⁰. The structural basis by which two heterodimers associate remains unexplored, but the present Ku–DNA structure seems to represent one half site of the end-to-end complex, albeit with a truncated DNA substrate (Figs 2b and 5a). The shape of the DNA-binding site on Ku seems well designed to structurally support but not obscure DNA ends. The combination of a broad base and narrow bridge allows Ku to cradle two turns of DNA while exposing much of the DNA surface area. This is

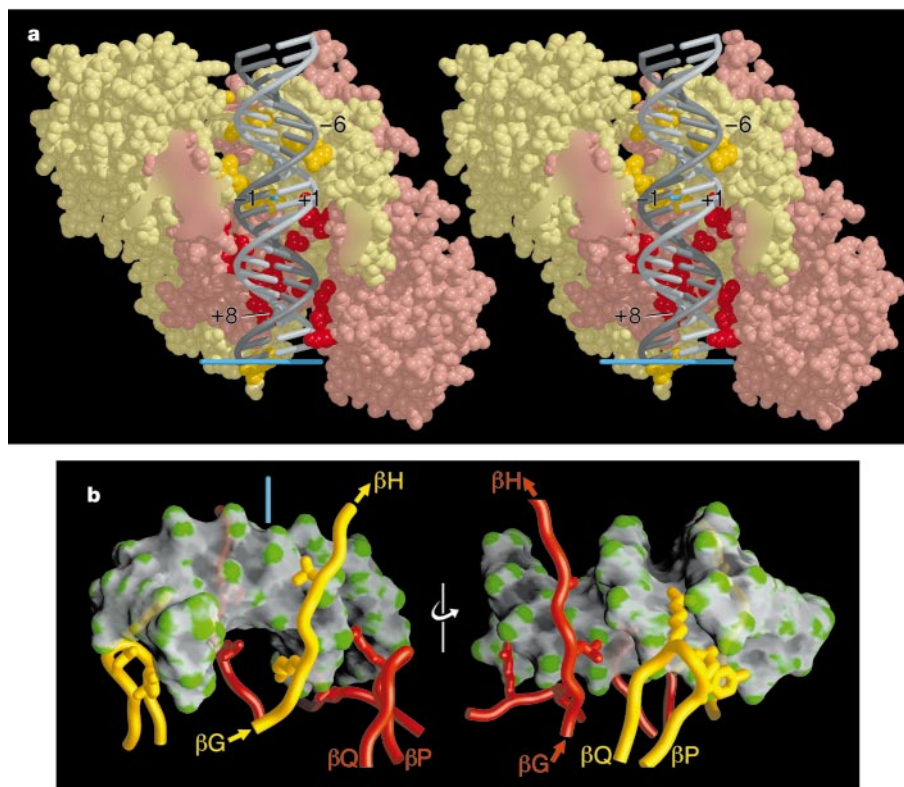


Figure 5 DNA-binding groove. **a**, Space-filling model of Ku. The bridge region has been sliced off to reveal the DNA. Twenty-two base pairs of B DNA were fitted to the crystallographically observed duplex, extending from the –11 level to the +11 level at bottom. The extent of the experimentally observed DNA is indicated (–6 and +8). Ku70 is coloured light red and Ku80 light orange. Protein atoms in bold colours constrain the phase of the DNA in the binding groove, as they lie inside a cylindrical ‘Van der Waals volume’, which contains the DNA (see text for details). This model may represent one half

site of an end-to-end tetrameric complex, with the broken DNA end at the bottom. The blue line drawn at the bottom shows one possible position for a dyad axis that would generate the other half site. **b**, Two views of the molecular surface of DNA, calculated from the crystal structure and coloured according to the degree of curvature at the surface (green, convex; grey, concave) using the program GRASP⁴⁴. Regions of Ku70 (red) and Ku80 (orange) that fit to major and minor groove contours are drawn as a backbone tube with selected side chains.

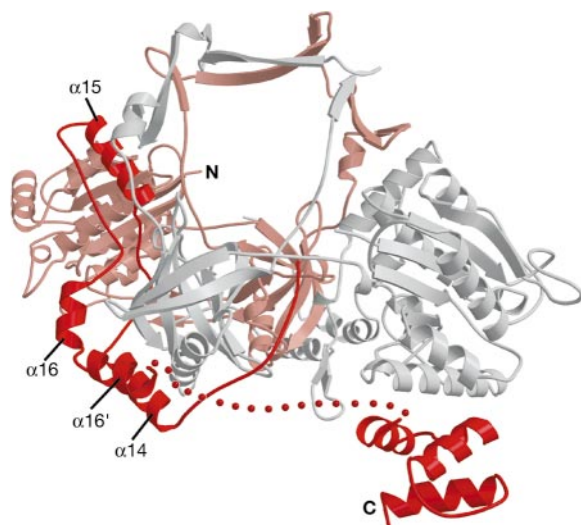


Figure 6 C-terminal DNA-binding SAP domain of Ku70. The ribbon diagram is of Ku crystallized in the absence of DNA. Ku70 is coloured light red and Ku80 grey. The C terminus of Ku70 is drawn in dark red from the C-terminal arm through to the terminus (labelled C). The dotted line indicates a disordered linker region between residues 539–558.

highlighted by a calculation relating to the model in Fig. 5a: of the 11 bp of DNA extending from the molecular dyad (+1 bp level) to the broken end (+11 level), ~70% of total DNA surface area remains exposed when bound to Ku (calculated with a probe radius of 1.4 Å). Biochemical and structural studies are required for further exploration of this issue, but overall it seems that alignment-based end joining will require the proper combination of DNA–protein interactions and DNA-surface exposure.

There have been reports of Ku binding tightly to specific DNA sequences embedded in circular plasmids, which is clearly precluded by the ring topology of Ku. The apparent specificity of Ku binding to a sequence element in mouse mammary tumour virus³⁴ is probably explained by the non-random (predominantly purines on one strand) DNA sequence having propensity towards a hinged configuration with triplex and single-stranded regions³⁵. Ku may bind and stabilize transient hairpin structures in such non-B-form DNA.

Major and minor grooves

Ku binds the 14-bp duplex in a sequence-independent manner through a limited set of interactions with the sugar-phosphate backbone. The absence of even a single interaction with a DNA base explains the lack of sequence specificity in end binding²². The closest approach to bases is from loop βP–βQ of the β-barrels, which enters the minor groove and positions an arginine residue (Ku70, Arg 403; Ku80, Arg 400) 4–5 Å from purine N3 groups at DNA levels +2 to +4 (Ku70) and –2 to –4 (Ku80) (Fig. 5b).

When viewed in projection along the DNA helix, the hole in the Ku ring appears significantly narrower than the diameter of DNA (Fig. 2a), owing to protein residues lying in the major and minor grooves, of which the βP–βQ loop is one example. Protein regions fit into the major groove at levels +1 to +4 (Ku70) and –1 to –4 (Ku80), where the polypeptide rings contribute to the pillar region (Figs 2c and 5b), although few of these residues actually contact DNA. Protein residues that enter the major and minor grooves were identified on a more quantitative basis: the (least-squares) axis of the crystallographically observed duplex was calculated and a radius of 11.5 Å applied to define a cylindrical Van der Waals volume. This volume should be free of protein atoms if DNA is to translate directly through the Ku ring, as seems to be the case for replication

processivity factors²⁷. Protein atoms whose own Van der Waals radii lie >0.5 Å within this volume were identified, and these are highlighted in Fig. 5a. The analysis shows that although there are no obstructions to DNA at either end of the binding groove, many residues (a roughly symmetrical set from each subunit) do lie substantially within the exclusion volume. Overall, these features will constrain DNA movement to a helical path through the Ku ring.

Role of Ku in repair complex assembly

Studies of NHEJ in *Xenopus* and mammalian systems have demonstrated a marked capacity of the repair complex to maintain the alignment of DNA termini (for example, for joining of blunt ends) and, when complementary base pairs are available, to position the junctional overlap on the basis of very short (1–4 bp) homologies, while allowing fill-in DNA synthesis before ligation^{9,17,18}. This capability is lost in Ku-deficient cells⁹. The ability of Ku to confine DNA movement to a helical path while reducing other significant degrees of freedom may be important in assisting the alignment of DNA ends. Ku provides a surface that cradles DNA with contours that complement a B-form helix, and this will help prevent DNA fraying and other low-frequency motions that would tend to mediate against the forces available to base pairing and stacking interactions. At the same time, in the alignment complex, DNA ends could search for their preferred overlap setting by moving along their natural helical paths.

These ideas focus attention on the geometry of the repair complex. If two Ku heterodimers do in fact form the core of such a complex then its dimensions can be crudely estimated by modelling using the half-site complex drawn in Fig. 5a. The heterodimers would presumably come together around a dyad axis, and the best fit is found when the molecular dyads of individual heterodimers are 60–90° apart, as their DNA-binding surfaces are then co-planar (Fig. 5a). In the resulting complex, there would be a distance of 20–24 bp of B DNA between the molecular dyads, ~12 bp of which would be accessible to processing enzymes within an access region of 50–60 Å exposed between the bridges of the two heterodimers. The effective area of access could be made even larger if the complex as a whole translocated along DNA, exposing a repair window.

Finally, the role of DNA-PK_{cs} in NHEJ is unclear, but involves its recruitment to the DNA end by Ku and activation of the protein kinase function²². Recent photo-crosslinking studies of the recruitment process³⁶ suggest that before DNA-PK_{cs} binding Ku resides at the DNA terminus with DNA extending to about the +10 level (Fig. 5a; compare with ref. 36 Fig. 9). On DNA-PK_{cs} binding, Ku shifts ~10 bp inwards and DNA-PK_{cs} itself adopts a position at the terminus. This accords with the predicted position of the DNA-PK_{cs} recruitment domain (C-terminal 19K of Ku80; ref. 25) being proximal to the free DNA end—the final ordered residue on Ku80, Lys 545, is only ~10 Å from DNA at level +8. When bound to DNA in this way, DNA-PK_{cs} may interfere with end processing, such that regulatory phosphorylation of DNA-PK_{cs} target proteins is required to promote dissociation of the complex and allow resumption of the joining process, as suggested previously^{10,37}.

Fate of Ku trapped on DNA

It is unclear how many molecules of Ku thread onto a DSB before the joining reaction is accomplished, but even a single, trapped heterodimer would seem to present a challenge, for example to progression of the replication fork. If there is an active mechanism for unloading Ku, it will have to contend with an extensive and stable subunit interface. Unloading could be achieved in principle by an ATP-dependent mechanism of the type used by the *Escherichia coli* γ-complex to unload processivity factor³⁸, but this is a reversible system that would tend to load, not unload, Ku. A proteolytic mechanism, on the other hand, would have the advantage of irreversibility, and the bridge regions are obvious sites at which to initiate such a process (Fig. 3c).

Ku has no sequence specificity for DNA ends, but shows apparent specificity in the form of pause sites when translocating along DNA, according to footprinting experiments³². A possible explanation is suggested by the finding that the 5K C terminus of Ku70 comprises a DNA-binding SAP domain²⁹ (Fig. 6). The DNA-binding capability of this region (Ku70 residues 536–609) has been demonstrated by southwestern blotting³⁹, but is too weak to detect in a gel-shift assay. In the crystals of DNA-free Ku, the SAP domain (residues 559–609) follows a 23-residue disordered linker, and packs against the base of the Ku80 α/β domain by means of a small but well-ordered interface, so that the recognition helix (residues 596–609) of the helix-turn-helix motif is exposed to solvent. In the Ku–DNA crystals the domain is disordered, and appears to have moved into a solvent void that also contains the disordered 3 bp of the G-rich duplex stem (Fig. 1b). The presence of a DNA-binding domain on Ku in addition to the ring is puzzling. The SAP domain is unlikely to be directly involved in the joining reaction as it is located distal to the free DNA end; even with the disordered linker region fully extended it would lie ~ 40 Å from the DNA end. Instead, the low affinity for DNA binding³⁹ suggests that this domain may exert its function when Ku has encircled DNA, perhaps by providing a barrier to inward movement of Ku away from the junction either during or after repair, or to cause pausing of Ku at specific DNA sequences. □

Methods

Protein expression and purification

Full-length Ku heterodimer was prepared by co-expression of Ku70 and Ku80 in insect cells from individual baculoviruses (Bac-to-Bac, Gibco). The heterodimer was probed with trypsin, releasing the 19K C-terminal domain of Ku80 (trypsin cleaves after Lys 565). An additional baculovirus was prepared for co-expression of truncated Ku80 (residues 1–579) with full-length Ku70; neither subunit bears a tag for purification. Insect cells were collected 48 h after infection and protein was purified by ammonium sulphate fractionation (25–75%) followed by chromatography through diethylaminoethyl-sepharose onto a heparin-sepharose column, DNA (calf thymus)-sepharose, and a final step of size-exclusion chromatography (Superdex 200). Selenomethionine was incorporated into Ku by substituting standard insect-cell culture medium with Met/Cys-free medium plus 2 mM cysteine and 60 mg l⁻¹ L-selenomethionine. The protocol achieved $\sim 70\%$ incorporation of selenomethionine.

Crystallization and structure determination

For crystallization of DNA-bound Ku, HPLC-purified 34- and 21-residue oligonucleotides (OligosEtc) were annealed and added in 1.5-fold molar excess to Ku heterodimer. The complex was crystallized at 22 °C by the hanging-drop method by adding 1 μ l of the 18 mg ml⁻¹ macromolecular solution to 1 μ l of a solution containing 26% polyethylene glycol (PEG) 350 monomethyl ether, 50 mM Tris-HCl (pH 8.0) and 5 mM MgCl₂. For diffraction studies, crystals were transferred to mother liquor solution containing an additional 2.5% glycerol, and flash-frozen in liquid propane. Diffraction data from a single Ku–DNA crystal were measured at beamline X9A of the National Synchrotron Light Source (NSLS), and data from crystals containing iodo- and bromo-derivatized oligonucleotides were measured at beamline F-2 of the Cornell High Energy Synchrotron Source (CHESS). Data were processed with programs DENZO and SCALEPACK⁴⁰.

For the crystallization of DNA-free Ku, 2 μ l of a 14 mg ml⁻¹ protein solution containing 900 mM sarcosine was mixed with 1 μ l of a reservoir solution of 15% PEG 4000, 900 mM sarcosine, 140 mM MgCl₂ and 100 mM Tris-HCl (pH 8.5). Before flash-freezing, crystals were cryo-protected by transfer to mother liquor solution containing 15% glycerol. The crystals grow to dimensions of $20 \times 70 \times 200$ μ m in two days; they appear to be single but in most cases are in fact multiple crystals. Only 1–2% of crystals were suitable for data collection, and this precluded structure determination by MIR. MAD data were collected from a single frozen crystal at beamline X25 of the NSLS. The crystals are radiation sensitive, so inverse-beam data collection was restricted to the anomalous peak wavelength, and 90° of data were collected at the other wavelengths. Data were processed as before, and all MAD data analysis was done with the program SOLVE⁴¹ using data between 25.0 and 3.5 Å resolution. SOLVE found 27 out of the possible 29 selenium sites and, after their refinement, gave a mean figure of merit of 0.32 (0.18 for the highest resolution bin). Density modification using the SOLVE package yielded an initial electron density map of excellent quality, albeit medium resolution, and a near-complete model was built aided by the selenium positions and the equivalence of Ku70 and Ku80. The refinement of atomic positions and temperature factors, together with a bulk solvent correction, using the program CNS⁴² reduced the *R*-factor to a final value of 22.2% (*R*_{free} of 28.4%) for data between 20.0 and 2.7 Å (with no amplitude cut-off). The model comprises 8,590 protein atoms and 43 water molecules, with no Ramachandran violations. The following residues were not modelled owing to weak or no associated electron density: Ku70 1–34, 224–230, 539–558 and Ku80 1–5, 170–181, 190–191, 324–326, 543–565.

The structure of DNA-bound Ku was determined by molecular replacement with the

program AMORE⁴³ using the model described above. A translation function calculated with data between 8.0 and 4.5 Å resolution gave a single, unambiguous (19 σ) peak. The positioned model was improved by rigid-body and all-atom refinement, followed by manual building of the DNA. DNA sequence assignment was aided by data from three crystals containing DNA labelled with iodo-U/C at 11 positions (crystal 11/13 in Table 1 of Supplementary Information), iodo-U/C at 11 positions (12/13), or bromo-U/C at 23 positions (B1/B2). Additional rounds of positional and individual *B*-factor refinement to 2.5 Å resolution with the program CNS⁴² reduced the *R*-factor to a final value of 21.9% (*R*_{free} of 27.5%) for data between 25.0 and 2.5 Å (with no amplitude cut-off). The final model is complete with the exception of the disordered Ku70 residues 1–33, 223–230, 533–609 (including the SAP domain), and Ku80 residues 1–5, 171–180, 546–565, plus the disordered final 3 bp of the G-rich stem that occupy a solvent space in the crystal. The average *B* factor for all protein atoms is 35 Å²; for the 14-bp duplex portion of DNA bound to Ku it is 81 Å², reflecting the paucity of protein–DNA contacts; and for the remaining 21 nucleotides it is 89 Å². The final model comprises 8,228 protein atoms, 993 DNA atoms and 255 water molecules, with no Ramachandran violations.

Received 30 May; accepted 29 June 2001.

- van Gent, D. C., Hoeijmakers, J. H. & Kanaar, R. Chromosomal stability and the DNA double-stranded break connection. *Nature Rev. Genet.* **2**, 196–206 (2001).
- Lieber, M. R. The biochemistry and biological significance of nonhomologous DNA end joining: an essential repair process in multicellular eukaryotes. *Genes Cells* **4**, 77–85 (1999).
- Critchlow, S. E. & Jackson, S. P. DNA end-joining: from yeast to man. *Trends Biochem. Sci.* **23**, 394–398 (1998).
- Haber, J. E. Partners and pathways: repairing a double-strand break. *Trends Genet.* **16**, 259–264 (2000).
- Gottlieb, T. M. & Jackson, S. P. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131–142 (1993).
- Smith, G. C. & Jackson, S. P. The DNA-dependent protein kinase. *Genes Dev.* **13**, 916–934 (1999).
- Boulton, S. J. & Jackson, S. P. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15**, 5093–5103 (1996).
- Liang, F. & Jasin, M. Ku80-deficient cells exhibit excess degradation of extrachromosomal DNA. *J. Biol. Chem.* **271**, 14405–14411 (1996).
- Feldmann, E., Schmiemann, V., Goedecke, W., Reichenberger, S. & Pfeiffer, P. DNA double-strand break repair in cell-free extracts from Ku80-deficient cells: implications for Ku serving as an alignment factor in non-homologous DNA end joining. *Nucleic Acids Res.* **28**, 2585–2596 (2000).
- Chen, S. *et al.* Accurate *in vitro* end-joining of a DNA double-strand break with partially cohesive 3'-overhangs and 3'-phosphoglycolate termini: effect of Ku on repair fidelity. *J. Biol. Chem.* **276**, 24323–24330 (2001).
- Taccioli, G. E. *et al.* Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. *Science* **265**, 1442–1445 (1994).
- Zhu, C., Bogue, M. A., Lim, D. S., Hasty, P. & Roth, D. B. Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates. *Cell* **86**, 379–389 (1996).
- Nussenzweig, A., Sokol, K., Burgman, P., Li, L. & Li, G. C. Hypersensitivity of Ku80-deficient cell lines and mice to DNA damage: the effects of ionizing radiation on growth, survival, and development. *Proc. Natl Acad. Sci. USA* **94**, 13588–13593 (1997).
- Gu, Y., Jin, S., Gao, Y., Weaver, D. T. & Alt, F. W. Ku70-deficient embryonic stem cells have increased radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination. *Proc. Natl Acad. Sci. USA* **94**, 8076–8081 (1997).
- Difilippantonio, M. J. *et al.* DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature* **404**, 510–514 (2000).
- Ferguson, D. O. *et al.* The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations. *Proc. Natl Acad. Sci. USA* **97**, 6630–6633 (2000).
- Roth, D. B. & Wilson, J. H. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol. Cell. Biol.* **6**, 4295–4304 (1986).
- Thode, S., Schaefer, A., Pfeiffer, P. & Vielmetter, W. A novel pathway of DNA end-to-end joining. *Cell* **60**, 921–928 (1990).
- Cary, R. B. *et al.* DNA looping by Ku and the DNA-dependent protein kinase. *Proc. Natl Acad. Sci. USA* **94**, 4267–4272 (1997).
- Ramsden, D. A. & Gellert, M. Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J.* **17**, 609–614 (1998).
- Nick McElhinny, S. A., Snowden, C. M., McCarville, J. & Ramsden, D. A. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol. Cell. Biol.* **20**, 2996–3003 (2000).
- Dynan, W. S. & Yoo, S. Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Res.* **26**, 1551–1559 (1998).
- Baumann, P. & West, S. C. DNA end-joining catalyzed by human cell-free extracts. *Proc. Natl Acad. Sci. USA* **95**, 14066–14070 (1998).
- Paillard, S. & Strauss, F. Site-specific proteolytic cleavage of Ku protein bound to DNA. *Proteins* **15**, 330–337 (1993).
- Singleton, B. K., Torres-Arzayus, M. I., Rottinghaus, S. T., Taccioli, G. E. & Jeggo, P. A. The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. *Mol. Cell. Biol.* **19**, 3267–3277 (1999).
- Yoo, S., Kimzey, A. & Dynan, W. S. Photocross-linking of an oriented DNA repair complex. Ku bound at a single DNA end. *J. Biol. Chem.* **274**, 20034–20039 (1999).
- Kong, X.-P., Onrust, R., O'Donnell, M. & Kuriyan, J. Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* **69**, 425–437 (1992).
- Gell, D. & Jackson, S. P. Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res.* **27**, 3494–4502 (1999).
- Aravind, L. & Koonin, E. V. SAP—a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem. Sci.* **25**, 112–114 (2000).
- Ghosh, G., van Duyn, G., Ghosh, S. & Sigler, P. B. Structure of NF- κ B p50 homodimer bound to a κ B site. *Nature* **373**, 303–310 (1995).

31. Muller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. Structure of the NF-kappa B p50 homodimer bound to DNA. *Nature* **373**, 311–317 (1995).
32. de Vries, E., van Driel, W., Bergsma, W. G., Arnberg, A. C. & van der Vliet, P. C. HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *J. Mol. Biol.* **208**, 65–78 (1989).
33. Zhao, J., Wang, J., Chen, D. J., Peterson, S. R. & Trewheella, J. The solution structure of the DNA double-stranded break repair protein Ku and its complex with DNA: a neutron contrast variation study. *Biochemistry* **38**, 2152–2159 (1999).
34. Giffin, W. *et al.* Sequence-specific DNA binding by Ku autoantigen and its effects on transcription. *Nature* **380**, 265–268 (1996).
35. Htun, H. & Dahlberg, J. E. Topology and formation of triple-stranded H-DNA. *Science* **243**, 1571–1576 (1989).
36. Yoo, S. & Dynan, W. S. Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res.* **27**, 4679–4686 (1999).
37. Calsou, P. *et al.* The DNA-dependent protein kinase catalytic activity regulates DNA end processing by means of Ku entry into DNA. *J. Biol. Chem.* **274**, 7848–7856 (1999).
38. Stewart, J., Hingorani, M. M., Kelman, Z. & O'Donnell, M. Mechanism of β clamp opening by the δ subunit of *E. coli* DNA polymerase III holoenzyme. *J. Biol. Chem.* **276**, 19182–19189 (2001).
39. Chou, C. H., Wang, J., Knuth, M. W. & Reeves, W. H. Role of a major autoepitope in forming the DNA binding site of the p70 (Ku) antigen. *J. Exp. Med.* **175**, 1677–1684 (1992).
40. Otwinoski, W. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
41. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
42. Brünger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
43. CCP4. The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank M. Jasin and K. Mariani for critical reading of the manuscript; J. Gulbis for assistance with baculovirus production; L. Berman and M. Becker for use of synchrotron facilities at NSLS and C. Heaton at CHESS; and P. Jeffrey for help with synchrotron data collection. This work was supported by grants to J.G. from the NIH, HHMI and Pew Scholars Program in the Biomedical Sciences.

Correspondence and requests for materials should be addressed to J.G. (e-mail: jonathan@ximpact4.ski.mskcc.org). The atomic coordinates have been deposited in the Protein Data Bank under accession numbers 1JEQ and 1JEY.

History of trace gases in presolar diamonds inferred from ion-implantation experiments

A. P. Koscheev*, M. D. Gromov*, R. K. Mohapatra† & U. Ott‡

* Karpov Institute of Physical Chemistry, Vorontzovo Pole 10, 103064, Moscow, Russia

† Max-Planck-Institut für Chemie (Otto-Hahn-Institut), Becherweg 27, D-55128 Mainz, Germany

Diamond grains are the most abundant presolar grains found in primitive meteorites^{1–3}. They formed before the Solar System, and therefore provide a record of nuclear and chemical processes in stars and in the interstellar medium^{1–3}. Their origins are inferred from the unusual isotopic compositions of trace elements—mainly xenon^{1–4}—which suggest that they came from supernovae. But the exact nature of the sources has been enigmatic, as has the method by which noble gases were incorporated into the grains. One observation is that different isotopic components are released at different temperatures when the grains are heated, and it has been suggested that these components have different origins. Here we report results of a laboratory study that shows that ion implantation (previously suggested on other grounds^{5,6}) is a viable mechanism for trapping noble gases. Moreover, we find that ion implantation of a single isotopic composition can produce both low- and high-temperature release peaks from the same grains. We conclude that both isotopically normal and anomalous gases may have been implanted by multiple events separated in space and/or time, with thermal processing producing an apparent enrichment of the anomalous component in the high-temperature release peak. The previous assumption that the low- and high-temperature components were not correlated may therefore have led to an overestimate of the abundance of anomalous argon and krypton, while obscuring an enhancement of the light—in addition to the heavy—krypton isotopes.

Although nanodiamonds are nominally the most abundant of the presolar grains so far identified in meteorites, doubts have remained as to what fraction of them is truly presolar. This is primarily because of two features that set them apart from other identified presolar grains such as silicon carbide, graphite and corundum. First, the isotopic composition of the structural element (carbon) itself is unremarkable, and second, the size of the diamond grains is extremely small, only about 2.6 nm (ref. 7), corresponding to about 1,000 carbon atoms. As a consequence, trace elements on which the characterization depends, such as xenon, occur with a very low abundance—only about one atom per million diamond grains. In addition, besides isotopically unusual trace-element components such as ‘xenon-HL’, which is strongly enriched in the heaviest and lightest of the Xe isotopes, other components of noble gases are found in diamond samples that are rather normal in their isotopic composition⁴. One way to address the problem is by gaining information about the mechanisms of diamond formation and the introduction of noble-gas trace elements.

Several mechanisms have been proposed for the formation of presolar nanodiamonds, including chemical vapour deposition (CVD)^{5,7}, high-pressure shock metamorphism⁸, and irradiation of carbonaceous grains by either ultraviolet radiation or energetic particles^{9,10}. As for incorporation of noble gases into the diamonds, two hypotheses have been seriously considered: (1) condensation and entrapment of noble gases by diamond grains as they grew, and (2) ion implantation—indirect evidence seems to favour the

latter^{1,5–7}. Previous laboratory experiments have shown the possibility of noble-gas trapping during synthesis of shock-produced diamonds¹¹ and CVD diamonds¹², as well as during ion implantation into carbonaceous matter^{13–15} and into other minerals^{16–18}. However, with only a few recent exceptions^{19,20}, the chemical nature of the samples studied was very different from that of nanodiamonds. In addition, the noble-gas release protocol generally was too different from that employed in the study of meteoritic diamond grains to perform a meaningful comparison. In contrast, in the experiments we report here, a mixture of noble-gas ions was implanted into synthetic nanodiamonds (ultradispersed detonation diamonds²¹, UDD) with a size of about 4 nm, close to that of the meteoritic ones, and the implanted noble gases were analysed using a protocol similar to the one used in studies of the meteoritic diamonds⁴.

The amounts of He, Ar, Kr and Xe trapped in the irradiated UDD are large, of the order of 10% of the ion dose (total, about 1.5×10^{15} ions cm⁻²) with which they had been irradiated. Observed concentrations (of ¹³²Xe, for example: $\sim 2 \times 10^{-5}$ cm³ STP g⁻¹) are much larger than previously observed in synthetic carbonaceous matter¹³ or CVD diamonds¹², and, except for He, where they are similar, even exceed those observed in presolar diamonds⁴. This confirms ion implantation as a viable mechanism for introduction of noble gases into grains of nanodiamond. The most notable observation, however, concerns the thermal release profiles (Fig. 1). Although details seem to depend on exact conditions during extraction (Figs 1 and 2; ref. 22), the main feature of these profiles is their bimodal character, with main peaks (for Ar, Kr, Xe) in the temperature range 200–700 °C and above 1,000 °C. The situation closely resembles that for diamonds from the most primitive meteorites⁴ analysed under similar conditions (Fig. 1).

Measurements of the isotopic composition of Ar, Kr and Xe (Fig. 2) released during stepped pyrolysis show the gases to be

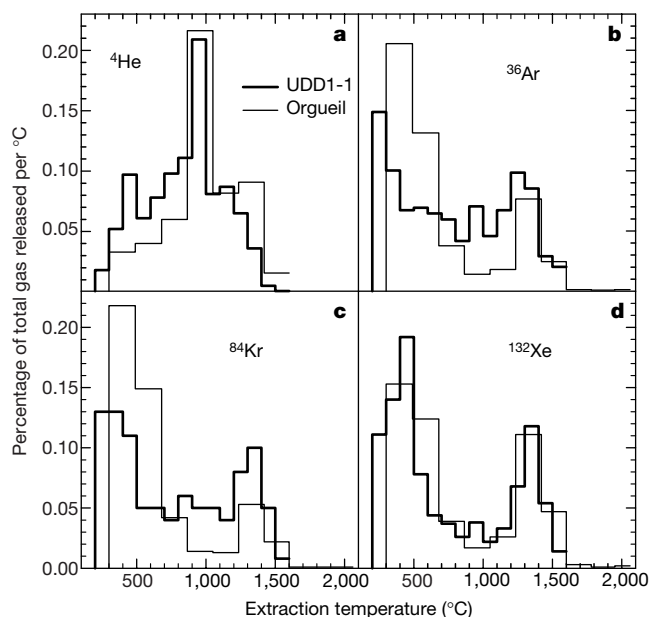


Figure 1 Thermal release patterns of noble gases from diamonds. We compare the release of He (a), Ar (b), Kr (c) and Xe (d) implanted into terrestrial nanodiamonds (sample UDD1-1: this work) with the release from presolar nanodiamonds extracted from the Orgueil meteorite⁴. The similarity is remarkable. Plotted in each case is the temperature-interval-normalized release (percentage of total gas released per °C). Data points for the lowest-temperature steps (in which little gas was released) are not shown because, with the exact previous temperature history of the diamonds unknown, we could not reliably calculate this quantity.

fractionated in a mass-dependent fashion, relative to atmospheric isotopic composition, favouring the heavy isotopes. For gas released in the low-temperature release peak, fractionation is small and needs to be studied in more detail (see also Methods section). However, observed fractionation increases with the temperature of the gas-release step, and, relative to the low-temperature release, gases released at high temperature are fractionated by ~ 0.8 (Xe), 1.0 (Kr) and 2.5% (Ar) per mass unit. Whereas the size of this fractionation effect is similar to that which has been observed in some previous implantation experiments^{14,16,17}, the simultaneous occurrence of a less fractionated (or possibly even unfractionated) component released in a separate low-temperature peak has so far been observed only in our experiments with nanodiamonds.

In the case of the presolar diamonds, the bimodal character of the release has commonly been attributed to the presence of different components of noble gases—namely P3 (with a ‘normal’ isotopic composition) released (only) in the low-temperature range, and HL (with ‘exotic’ isotopic composition) released (only) in the high-temperature range⁴ (we ignore here the minor P6 component). These components are thought to be uncorrelated, to have different origins and possibly to be located in different carrier grains⁴. But the results of our implantation experiments show that both the low- and high-temperature peaks may arise simultaneously in the same diamonds during a ‘one-component’ ion implantation event. As a consequence, if introduced in this manner, a high-temperature part of the P3 component must exist as well, with implications regarding the composition of isotopically unusual Xe-HL and the relative timing of the introduction of the different components. This is discussed in detail below.

Isotopically unusual Xe-HL is characterized by strong overabundances of the heavy isotopes (especially ^{134}Xe and ^{136}Xe , which are produced in the r-process of nucleosynthesis only) and the light ones (primarily p-process-only ^{124}Xe and ^{126}Xe) relative to those of intermediate mass such as s-process-only ^{130}Xe (ref. 4). However, the ratios $^{136}\text{Xe}/^{130}\text{Xe}$ and $^{124}\text{Xe}/^{130}\text{Xe}$ have never been found to exceed by more than ~ 2.2 and 1.8 times, respectively, the ratios observed in the solar wind. This has led either to the assumption that the Xe-HL component was directly produced with these ratios—which would require a special nucleosynthesis process—or to the conclusion that it originally had a more extreme composition but became mixed with normal-type xenon before incorpora-

tion into the diamonds. As pointed out above, an important consequence of the ion-implantation model is that within it a major amount of the (nominally low-temperature) P3 component must accompany Xe-HL during the high-temperature gas release. Given this observation, it is an obvious conclusion that possibly all (or at least a significant fraction) of the s-only ^{130}Xe observed in the high-temperature release actually belongs to P3 rather than to Xe-HL, and that Xe-HL indeed was implanted with an extreme composition, possibly without any s-process isotopes. This would validate an assumption inherent in most nucleosynthetic scenarios for production of Xe-HL (see, for example, refs 23–25), which generally try to reproduce an ^{130}Xe -free composition.

The observation that high- and low-temperature gas release peaks may be produced in the same implantation event also has implications for the timing of the noble-gas trapping processes. The following sequence of events seems to be required. The HL component must have been implanted into diamond grains in a first stage. As there is no low-temperature part of Xe-HL observed today, this then must have been lost, probably owing to elevated temperatures either during implantation or in some later episode. Such a loss may easily have taken place, given recent estimates for an equilibrium temperature of $\sim 1,000$ K for diamond grains in dusty envelopes surrounding stars²⁶—sufficient for loss of the low-temperature part of trapped Xe-HL gas without seriously affecting the high-temperature part. The incorporation of the P3 component (with low- and high-temperature part) must have happened at a different stage. This may have been ion implantation into the same population of grains after loss of the low-temperature part of Xe-HL gases and cooling of the grains. Alternatively, implantation might have happened (at any time) into a different subfraction of diamond grains, which later were intimately mixed with those bearing the Xe-HL gas.

How exact conditions during implantation (such as the dose²⁰) and possibly during gas extraction²² (compare also Figs 1 and 2) affect the degree of fractionation, as well as the distribution of implanted gases between low- and high-temperature release peaks, remains to be studied in detail. But another inference is that the isotopic fractionation of the high-temperature gases needs to be properly taken into account in deriving the abundance and composition of the anomalous components. It is clear in the case of xenon that the basic pattern derived for Xe-HL is not seriously affected, but the situation is different for the lighter noble gases

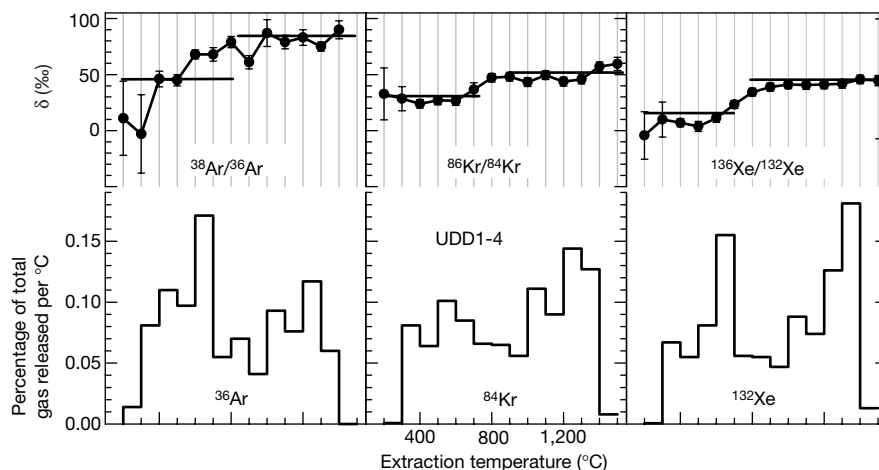


Figure 2 Thermal release and isotopic variation of implanted Ar, Kr and Xe. In each panel, we show the release from terrestrial nanodiamond sample UDD1-4 (bottom part) and the variation observed for isotopic ratios $^{38}\text{Ar}/^{36}\text{Ar}$, $^{86}\text{Kr}/^{84}\text{Kr}$ and $^{136}\text{Xe}/^{132}\text{Xe}$ (top part). Isotopic ratios are shown as deviations (in per mil) from atmospheric ratios. As in Fig. 1, the thermal release shown is temperature-interval-normalized, and no data points are

shown for the first release step. Like UDD1-1 (Fig. 1), UDD1-4 shows a bimodal release pattern. Isotopic fractionation at low temperature is modest, and may be explained by the experimental set-up (see text and Methods). Isotopic fractionation of gas released at high temperature is significantly higher.

where fractionation is larger and where the observed isotopic anomalies are more modest⁴. A major puzzle in the study of noble gases in presolar diamond has always been that, in the case of xenon, overabundances in the light isotopes ('Xe-L') accompany those in the heavy ones ('Xe-H'), but that—in the case of krypton—'Kr-H' is accompanied by what seems a deficit rather than an overabundance of the light isotopes ⁷⁸Kr and ⁸⁰Kr relative to s-only ⁸²Kr (refs 4, 27). A first inspection of the changes afforded by taking isotopic fractionation into account²⁸ indicates that this puzzle may go away and that Kr-L does exist after all, while there may be little (if any) argon associated with anomalous krypton and xenon. □

Methods

The nanodiamonds—obtained from the Scientific Research Institute of Technical Physics (Snezhinsk, Russia)—were produced as ultradispersed detonation diamonds²¹ (UDD) by detonation synthesis from explosives, with a mean size of ~4 nm. For the implantation an ordinary ion gauge set-up was used, with an axial tungsten cathode surrounded by accelerating grid and cylindrical tantalum collector. Implantation was from a flow of noble gases (estimated ion intensities, He:Ar:Kr:Xe ≈ 150:50:1:1) into a ~3 × 10⁻⁴ g cm⁻² nanodiamond layer deposited from a colloidal suspension onto part of the inner surface of the collector. With an energy of ~700 eV, the range of the ions in diamond is of the order of the grain radius, according to calculations using the TRIM code²⁹, except for He with a somewhat larger range. After implantation the diamonds were scraped off the collector. Gases were released by vacuum pyrolysis and analysed for the abundance and isotopic composition of noble gases by standard noble-gas mass spectrometry^{4,30}.

For the implantation, gases were introduced from a reservoir at about one bar pressure into the diffusion-governed regime of the ion source at 5 × 10⁻⁴ mbar. Further calibrations will be required in order to decide whether the small isotopic fractionation (relative to atmospheric isotopic composition) observed for the gases in the low-temperature peak (Fig. 2)—which closely follows an *m*^{1/2} fractionation law, where *m* is atomic mass—is associated with flow conditions or with the implantation process itself.

Received 5 February; accepted 24 May 2001.

- Anders, E. & Zinner, E. Interstellar grains in primitive meteorites: diamond, silicon carbide, and graphite. *Meteoritics* **28**, 490–514 (1993).
- Ott, U. Interstellar grains in meteorites. *Nature* **364**, 25–33 (1993).
- Zinner, E. in *Astrophysical Implications of the Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E.), 3–26 (American Institute of Physics, Woodbury, New York, 1997).
- Huss, G. R. & Lewis, R. S. Noble gases in presolar diamonds I: Three distinct components and their implications for diamond origins. *Meteoritics* **29**, 791–810 (1994).
- Lewis, R. S., Tang, M., Wacker, J. G., Anders, E. & Steel, E. Interstellar diamonds in meteorites. *Nature* **326**, 160–162 (1987).
- Verchovsky, A. B. et al. C, N, and noble gas isotopes in grain size separates of presolar diamonds from Efremovka. *Science* **281**, 1165–1168 (1998).
- Daulton, T. L., Eisenhour, D. D., Bernatowicz, T. J., Lewis, R. S. & Buseck, P. R. Genesis of presolar diamonds: comparative high-resolution transmission electron microscopy study of meteoritic and terrestrial nano-diamonds. *Geochim. Cosmochim. Acta* **60**, 4853–4872 (1996).
- Tielens, A. G. G. M., Seab, C. G., Hollenbach, D. J. & McKee, C. F. Shock processing of interstellar dust: diamonds in the sky. *Astrophys. J.* **319**, L109–L113 (1987).
- Nuth, J. A. III & Allen, J. E. Jr Supernovae as sources of interstellar diamonds. *Astrophys. Space Sci.* **196**, 117–123 (1992).
- Ozima, M. & Mochizuki, K. Origin of nanodiamonds in primitive chondrites: (1) Theory. *Meteoritics* **28**, 416–417 (1993).
- Matsuda, J., Kusumi, A., Yajima, H. & Syono, Y. Noble gas studies in diamonds synthesized by shock loading in the laboratory and their implications on the origin of diamonds in ureilites. *Geochim. Cosmochim. Acta* **59**, 4939–4949 (1995).
- Matsuda, J., Fukunaga, K. & Ito, K. Noble gas studies in vapor-growth diamonds: comparison with shock-produced diamonds and the origin of diamonds in ureilites. *Geochim. Cosmochim. Acta* **55**, 2011–2023 (1991).
- Frick, U., Mack, R. & Chang, S. Noble gas trapping and fractionation during synthesis of carbonaceous matter. *Proc. Lunar. Planet. Sci. Conf.* **X**, 1961–1973 (1979).
- Bernatowicz, T. J. & Fahey, A. J. Xe isotopic fractionation in a cathodeless glow discharge. *Geochim. Cosmochim. Acta* **50**, 445–452 (1986).
- Weigel, A. et al. Noble gas systematics in planetary atmospheres: Simulation of fractionation mechanisms using ion implantation. *Lunar Planet. Sci. [CD-ROM] XXIX*, abstr. 1900 (1998).
- Bernatowicz, T. J. & Hagee, B. E. Isotopic fractionation of Kr and Xe implanted in solids at very low energies. *Geochim. Cosmochim. Acta* **51**, 1599–1611 (1987).
- Pongani, K. V., Graf, T. & Marti, K. Isotopic fractionation in low-energy ion implantation. *J. Geophys. Res.* **102**, 19335–19343 (1997).
- Futagami, T., Ozima, M., Nagai, S. & Aoki, Y. Experiments on thermal release of implanted noble gases from minerals and their implications for noble gases in lunar soil grains. *Geochim. Cosmochim. Acta* **57**, 3177–3194 (1993).
- Koscheev, A. P., Gromov, M. D., Herrmann, S. & Ott, U. Trapping and isotope fractionation of noble gases in synthetic analog of presolar diamond grains. *Lunar Planet. Sci. [CD-ROM] XXXI*, abstr. 1551 (2000).
- Verchovsky, A. B., Wright, I. P., Fisenko, A. V., Semjonova, L. F. & Pillinger, C. T. Ion implantation into presolar diamonds: experimental simulation. *J. Conf. Abstr. [CD-ROM] 5*, abstr. 1050 (2000).

- Lyamkin, A. I. et al. Production of diamonds from explosives. *Sov. Phys. Dokl.* **33**, 705–706 (1988).
- Koscheev, A. P. & Ott, U. Mechanism of noble gas release during pyrolysis of nanodiamond grains. *Meteorit. Planet. Sci.* **35**, A92 (2000).
- Clayton, D. D. Origin of heavy xenon in meteoritic diamonds. *Astrophys. J.* **340**, 613–619 (1989).
- Howard, W. M., Meyer, B. S. & Clayton, D. D. Heavy-element abundances from a neutron burst that produces Xe-H. *Meteoritics* **27**, 404–412 (1992).
- Ott, U. Interstellar diamond xenon and timescales of supernova ejecta. *Astrophys. J.* **463**, 344–348 (1996).
- Guillois, O., Ledoux, G. & Reynaud, C. Diamond infrared emission bands in circumstellar media. *Astrophys. J.* **521**, L133–L136 (1999).
- Frick, U. Anomalous krypton in the Allende meteorite. *Proc. Lunar Sci. Conf. VIII*, 273–292 (1977).
- Huss, G. R., Ott, U. & Koscheev, A. P. Implications of ion-implantation experiments for understanding noble gases in presolar diamonds. *Meteorit. Planet. Sci.* **35**, A79–A80 (2000).
- Ziegler, J. F., Biersack, J. P. & Littmark, U. *The Stopping Range of Ions in Solids* (Pergamon, Tarrytown, New York, 1985).
- Schelhaas, N., Ott, U. & Begemann, F. Trapped noble gases in unequilibrated ordinary chondrites. *Geochim. Cosmochim. Acta* **54**, 2869–2882 (1990).

Acknowledgements

We thank S. Herrmann for assisting with the noble-gas measurements, G.R. Huss and A.B. Verchovsky for discussions, and R.S. Lewis and A.P. Meshik for comments on the manuscript. This work was partially supported by the Russian Foundation of Basic Science and by the German DFG.

Correspondence and requests for materials should be addressed to U.O. (e-mail: ott@mpch-mainz.mpg.de).

Two-dimensional imaging of electronic wavefunctions in carbon nanotubes

Serge G. Lemay*, Jorg W. Janssen*, Michiel van den Hout*, Maarten Mooij*, Michael J. Bronikowski†, Peter A. Willis†, Richard E. Smalley†, Leo P. Kouwenhoven* & Cees Dekker*

* Department of Applied Physics and DIMES, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

† Center for Nanoscale Science and Technology, Rice Quantum Institute, Departments of Chemistry and Physics, MS-100, Rice University, PO Box 1892, Houston, Texas 77251, USA

The drive towards the development of molecular electronics is placing increasing demands on the level of control that must be exerted on the electronic structure of materials. Proposed device architectures ultimately rely on tuning the interactions between individual electronic states, which amounts to controlling the detailed spatial structure of the electronic wavefunctions in the constituent molecules^{1,2}. Few experimental tools are available to probe this spatial structure directly, and the shapes of molecular wavefunctions are usually only known from theoretical investigations. Here we present scanning tunnelling spectroscopy measurements of the two-dimensional structure of individual wavefunctions in metallic single-walled carbon nanotubes; these measurements reveal spatial patterns that can be directly understood from the electronic structure of a single graphite sheet, and which represent an elegant illustration of Bloch's theorem³ at the level of individual wavefunctions. We also observe energy-dependent interference patterns in the wavefunctions and exploit these to directly measure the linear electronic dispersion relation of the metallic single-walled carbon nanotube.

We deposited nanotubes from a sonicated dichloroethane suspension⁴ onto atomically flat Au(111) surfaces. To reduce disorder in the observed electronic band structure (J.W.J., S.G.L.,

L.P.K. and C.D., unpublished results), we used high-purity carbon nanotubes that had been catalytically grown in high-pressure carbon monoxide⁵. To increase the electronic energy-level spacing, individual metallic single-walled nanotubes (SWNTs) were shortened to less than 40 nm by applying locally a short bias pulse of 6.5 V between the tip of the scanning tunnelling microscope (STM) and the sample⁶ (Fig. 1a). Topography (Fig. 1b) and scanning tunnelling spectroscopy (STS) measurements were then performed as a function of position on a two-dimensional grid with sub-ångström resolution.

The STS technique measures the tunnelling differential conductance (dI/dV) between the STM tip and the sample as a function of sample bias V , where dI/dV is to a good approximation proportional to the local density of electronic states (LDOS) of the sample⁷. For a system described by discrete electronic wavefunctions $\psi_j(\mathbf{r})$, the measured STS signal is thus given by

$$\frac{dI}{dV}(V, \mathbf{r}) \propto \sum_{|eV - \epsilon_j| < \delta} |\psi_j(\mathbf{r})|^2 \quad (1)$$

where $\mathbf{r}$ is the position coordinate and δ is the experimental energy resolution. When δ is less than the energy level spacing $\epsilon_{j+1} - \epsilon_j$, the sum in equation (1) reduces to a single term, and a two-dimensional measurement of the LDOS at fixed energy ϵ_j corresponds to a spatial map of $|\psi_j|^2$.

Figure 1c shows a measurement of dI/dV versus V performed on a

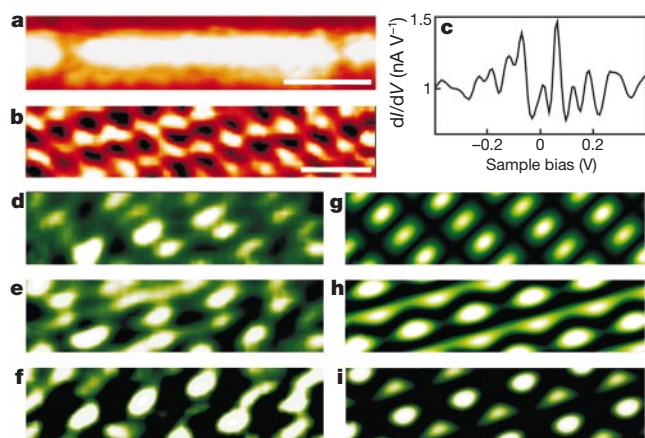


Figure 1 Energy-resolved images of individual molecular wavefunctions. **a**, Constant-current topographic image of a metallic SWNT that has been cut to a length of 34 nm. This nanotube has an apparent height of 1.1 nm and a chiral angle of 12°. The data were recorded in constant-current mode using a feedback current of 200 pA and a sample voltage of -200 mV. Scale bar, 10 nm. **b**, Constant-current topographic image of the atomic lattice of the shortened nanotube. Scale bar, 0.5 nm. **c**, Scanning tunnelling spectroscopy measurement. The vertical axis is approximately proportional to the LDOS; we attribute sharp peaks to discrete 'particle-in-a-box' electronic states. The data are a spatial average over the area shown in **b**. **d-f**, Energy-resolved images of the individual states at energies of -96, 30 and 96 meV, respectively, illustrating the variety in the appearance of individual wavefunctions. These dI/dV images are obtained using a lock-in technique. The periodicity of these images differs from that of the simultaneously acquired atomic lattice image shown in **b**. **g-i**, Calculated spatial maps of $|\psi_j(\mathbf{r})|^2$ based on equation (2). The characteristic features of each image are well reproduced. The calculation does not include a slow variation of the observed structure with position, described in detail in Fig. 3. All measurements were performed in an Omicron LT-STM operated at 4.6 K using Pt/Ir tips cut in air. Featureless $I-V$ characteristics were observed on the Au(111) substrate—both before and after STS measurements were performed on the SWNT. All data reported in this Letter are from the same sample (except Fig. 2d). Consistent results were obtained on a second shortened SWNT sample. In addition, similar two-dimensional patterns were observed on disordered ropes of SWNTs.

metallic SWNT. The presence of a series of sharp peaks in the LDOS indicates that the effective energy resolution δ is indeed smaller than the confinement-induced level spacing in our short SWNTs. Figure 1d-f shows three representative images of dI/dV versus position at fixed V . These correspond to spatial maps of different individual molecular wavefunctions $|\psi_j|^2$. All of the measured wavefunction images show a quasi-periodic pattern of spots with similar spacing, but each image also exhibits unique features such as stripes, alternating rows of bright and dim spots, or a rectangular supercell. The discrete wavefunctions of carbon nanotubes thus appear to display a variety of spatial patterns.

The wavefunction images differ from the image of the atomic lattice of the nanotube, shown in Fig. 1b. The nature of this difference is most clearly evident in a reciprocal-space representation. A SWNT can be thought of as a single plane of graphite rolled into a cylinder; the hexagonal lattice of carbon atoms is sketched in Fig. 2A, and the Fourier transform of an STM image of the atomic

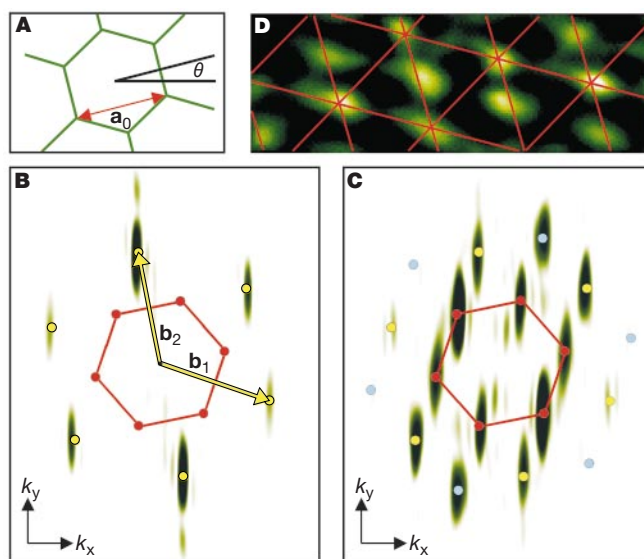


Figure 2 Comparison of the observed spatial structure with theory. The horizontal direction corresponds to the longitudinal axis of the SWNT. **A**, Real-space hexagonal lattice and definitions of the lattice spacing a_0 and of the chiral angle θ . **B**, Two-dimensional Fourier transform of a topographic image of the atomic lattice. Peaks appear at wavevectors corresponding to reciprocal lattice vectors $\mathbf{G} = m\mathbf{b}_1 + n\mathbf{b}_2$, where m, n are integers (filled yellow circles). The red hexagon indicates the border of the first Brillouin zone, and the red dots correspond to the calculated Fermi wavevectors $\pm\mathbf{k}_n^0$. The peaks in the data are elongated in the traverse (k_y) direction because the SWNT is intrinsically narrow in that direction. **C**, Two-dimensional Fourier transform of a wavefunction image at a sample bias $V = 64$ mV. As per equation (2), the wavefunction $\psi_j(\mathbf{r})$ is dominated by six Fourier components with wavevectors corresponding to the corners of the first Brillouin zone (red dots). Additional Fourier components corresponding to the second harmonics of these six fundamental components also appear (yellow and blue dots) because the experiment is sensitive to $|\psi_j(\mathbf{r})|^2$ instead of $\psi_j(\mathbf{r})$. For clarity, the high-intensity peak at $\mathbf{k} = 0$ was removed from the Fourier transforms in **B** and **C**.

D, Relation between a measured wavefunction image and the calculated Fourier components. Each set of parallel lines represents the wavefronts of one of the 'fundamental' Fourier components $\pm\mathbf{k}_n^0$ represented by red dots in **C**. The spacing and orientation of these wavefronts were determined from a topographic measurement of the atomic lattice; only the phase of each wave was adjusted to coincide with the position of the bright spots in the energy-resolved image. Some spots appear between crossing red lines. These are due to the second harmonics corresponding to the yellow and blue dots in **C**, and are present because we measure $|\psi_j(\mathbf{r})|^2$. Image **D** was obtained on a SWNT rope and is shown here because it is relatively wide in the transverse (y) direction, thus providing a clearer real-space illustration than our measurements on individual SWNTs.

lattice is shown in Fig. 2B. As expected, the Fourier transform exhibits bright spots at wavevectors $\mathbf{k}$ corresponding to the reciprocal lattice vectors $\mathbf{G}$. The Fourier transform of a wavefunction image, shown in Fig. 2C, shows new peaks at smaller wavevectors. In particular, the additional Fourier components in the wavefunction images have wavevectors that coincide with the corners of the first Brillouin zone (represented by a red hexagon in Fig. 2B, C).

Nanotube wavefunction images have been calculated before^{8,9}, but not yet experimentally verified. One-dimensional STS line scans have been reported for the special case of an 'armchair' nanotube¹⁰, and were interpreted in terms of a simple one-dimensional particle-in-a-box model. We now show that the full two-dimensional structure of measured wavefunctions as reported in our work can be understood, for arbitrary chirality, from a careful examination of the electronic structure of a single plane of graphite.

Band-structure calculations indicate that the Fermi surface of a plane of graphite consists of only six discrete points corresponding to the corners of the hexagonal first Brillouin zone. We label these wavevectors $\pm \mathbf{k}_n^0$, where $n = -1, 0, 1$. A wavefunction $\psi_j(\mathbf{r})$ at the Fermi level of a finite-sized SWNT is a superposition of Bloch waves, and can be written in the form⁹ $\psi_j(\mathbf{r}) = 2\text{Re}[u_{j,n}(\mathbf{r}) \exp(i\mathbf{k}_n^0 \cdot \mathbf{r})]$. Here $u_{j,n}(\mathbf{r})$ is a function with the periodicity of the atomic lattice. It follows that the Fourier transform of $\psi_j(\mathbf{r})$ contains Fourier components with wavevectors $\pm \mathbf{k}_n^0 + \{\mathbf{G}\}$, where $\{\mathbf{G}\}$ is the set of all

reciprocal lattice vectors. But not all of these Fourier components contribute to the measured images. Because Fourier components with larger $|\mathbf{k}|$ decay more rapidly with distance from the SWNT surface, STS images are dominated by the Fourier components with the lowest wavevectors⁷. The measured wavefunctions near the Fermi level are therefore predicted to have the form:

$$\psi_j(\mathbf{r}) = \sum_{n=-1}^1 (\phi_{j,n} e^{i\mathbf{k}_n^0 \cdot \mathbf{r}} + \phi_{j,n}^* e^{-i\mathbf{k}_n^0 \cdot \mathbf{r}}) \quad (2)$$

Equation (2) successfully describes the wavefunction images we observe, as demonstrated by the coincidence of the experimental peaks with the calculated dots appearing in the Fourier transform (Fig. 2C). Figure 2D further illustrates in real space the relationship between the Fourier components $\exp(i\mathbf{k}_n^0 \cdot \mathbf{r})$ and the periodic structures observed in the measured wavefunctions. It displays a wavefunction image on which lines have been superimposed to represent the lines of constant phase of each of these Fourier components. The predicted orientation and wavelength clearly match the rows of peaks observed in the data. Finally, Fig. 1g–i illustrates that the detailed features of the measured wavefunctions can be reproduced by equation (2): selecting appropriate values of $\phi_{j,n}$ yields a variety of spatial patterns that closely matches the experimental results shown in Fig. 1d–f. Slight differences between experimental (Fig. 1d–f) and calculated (Fig. 1g–i) images can be attributed to residual noise as well as to long-wavelength modulations observed in the measurements (discussed below).

We now turn to variations in the measured wavefunctions as a function of energy, from which we can extract information about the electronic dispersion relation of the SWNTs. The detailed spatial pattern of each wavefunction appears to exhibit slow variations with position that are not accounted for by equation (2). In order to clearly display these variations, we plot in Fig. 3b the LDOS at energy $E = -60$ meV as a function of position along the nanotube axis over a larger range of x than shown in Fig. 1. A slow, sinusoidal-like oscillation is found in addition to the rapid Fermi-wavelength oscillations discussed above. The wavelength of this slow oscillation increases markedly with increasing energy, as can be seen from the plot of the LDOS shown in Fig. 3c.

These energy-dependent features allow us to verify some key features of the band structure of carbon nanotubes. The theoretical dispersion relation of a metallic SWNT is sketched in Fig. 3a; two linear one-dimensional bands cross at the charge-neutrality point E_0 . These two bands are orthogonal in a SWNT of infinite length, but in our finite-sized tubes they mix because the atomic structure at the tube ends generally has low symmetry (A. A. Maarouf, N. R. Wilson and C. L. Kane, manuscript in preparation). Each individual wavefunction $\psi_j(\mathbf{r})$ is thus a linear combination of two left- and two right-moving Bloch waves with wavevectors $\pm (\mathbf{k}_0^0 \pm \delta \mathbf{k}_j)$, where $\delta \mathbf{k}_j$ is related to the energy ϵ_j by:

$$\delta k_j = (E_0 - \epsilon_j)/\hbar v_F \quad (3)$$

Here v_F is the Fermi velocity. Superposing waves with slightly different wavevectors $\mathbf{k}_0^0 \pm \delta \mathbf{k}_j$ results in a beating pattern—that is, a rapid oscillation with wavevector $\mathbf{k}_0^0$ modulated by an envelope function with wavevector $\delta \mathbf{k}_j$. The rapid oscillation with wavevector $\mathbf{k}_0^0$ is responsible for the primary structure in our wavefunctions (Figs 1 and 2). The additional slow modulation corresponds to the beating pattern envelope. The oscillations observed here are thus the result of quantum-mechanical interference at the level of individual electronic wavefunctions. They are fundamentally different from those reported near step edges of metallic surfaces^{11,12}, where the interference pattern instead originates from a continuum of states with a fixed phase relation at the step edge.

Measuring δk_j as a function of ϵ_j provides us with the means to directly extract the dispersion relation near the Fermi level for this individual SWNT. The result is shown in Fig. 3d. The dispersion

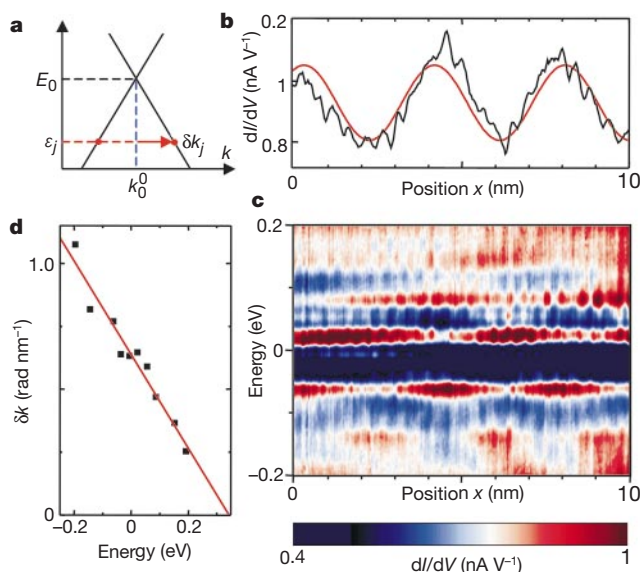


Figure 3 Determination of the dispersion relation from the energy dependence of wavefunction images. **a**, Sketch of the calculated one-dimensional band structure of a chiral metallic SWNT near the Fermi level. Two bands cross at the energy E_0 . Under our experimental conditions⁴, the Fermi energy E_F lies below E_0 . Individual wavefunctions are composed of Bloch waves with wavevectors $\pm (\mathbf{k}_0^0 \pm \delta \mathbf{k}_j)$, where $\delta \mathbf{k}_j$ is parallel to the axis of the SWNT and varies linearly with energy ϵ_j . **b**, Measurement of dI/dV versus position x along the longitudinal axis of the SWNT for $E = eV = -60$ meV. A low-wavevector modulation is apparent, which we associate with a beating pattern with wavevector $\delta \mathbf{k}_j$. The red line is a fit to $|\psi(x)|^2 = A \cos(2\delta k_j x + \varphi)$, where φ is an arbitrary phase and the factor of 2 appears because STS probes $|\psi(x)|^2$ rather than $\psi(x)$. **c**, Colour-scale plot of dI/dV versus electron energy E and position x along the longitudinal axis of the SWNT. Two types of features are observed: (1) an energy-independent modulation with wavevector $2\mathbf{k}_0^0$ that appears as vertical stripes (about 20 periods visible), and (2) a slow modulation whose wavevector varies with energy. For example, about 2.5 periods of this envelope are visible at $E = -0.06$ eV. **d**, Dispersion relation δk_j versus E_j . δk_j was obtained from a fit of each peak in **c**, as illustrated in **b**. The red line is a fit to the linear dispersion relation of equation (3), yielding values of $(8.2 \pm 0.7) \times 10^5$ m s⁻¹ for the Fermi velocity v_F and 0.34 ± 0.03 eV for the energy of the charge-neutrality point $E_0 - E_F$.

relation is linear, as predicted by equation (3). This measurement is, to our knowledge, the first direct experimental verification of this important property—on which (among other things) the application of Luttinger-liquid theory to carbon nanotubes is based^{13,14}.

Fitting the measured dispersion relation to equation (3) yields the value of the Fermi velocity in nanotubes, $v_F = (8.2 \pm 0.7) \times 10^5 \text{ m s}^{-1}$. In a tight-binding description, v_F is related to the π - π overlap energy γ_0 by $v_F = 3^{1/2} \gamma_0 a_0 / 2\hbar$, where a_0 is the atomic lattice spacing. Our measurement thus corresponds to a value of $\gamma_0 = 2.6 \pm 0.2 \text{ eV}$. For comparison, determinations of γ_0 based on the energies of van Hove singularities in STS and Raman spectroscopy measurements yield 2.5–2.9 and 2.6–3.0 eV, respectively¹⁵. Our determination of γ_0 based on a direct measurement of $E(k)$ performed near the Fermi level is therefore in agreement with those based on higher-energy structures. The fit also yields $E_0 - E_F = 0.34 \pm 0.03 \text{ eV}$ for the energy of the charge neutrality point ($\delta k = 0$), in agreement with previous estimates^{4,16}.

Analysis of the variation of the energy-level spacing $\epsilon_{j+1} - \epsilon_j$ and of the energy-dependence of the coefficients $\phi_{j,n}$ of our wavefunctions reveals departures from recent theoretical predictions (A. A. Maarouf, N. R. Wilson and C. L. Kane, manuscript in preparation), suggesting that the electronic boundary conditions at the tube ends are energy-dependent. Interference effects have recently been invoked to explain transport measurements on SWNTs¹⁷, but irregularities were observed that could also not be explained by assuming energy-independent boundary conditions. It is likely that harnessing these quantum-mechanical interference effects for practical devices will ultimately require the control of the atomic structure—and hence the electronic boundary conditions—at the nanotube ends. □

Received 4 June; accepted 13 July 2001.

1. Aviram, A. & Ratner, M. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277–283 (1974).
2. Joachim, C., Gimzewski, J. K. & Aviram, A. Electronics using hybrid-molecular and mono-molecular devices. *Nature* **408**, 541–548 (2000).
3. Bloch, F. Über die quantenmechanik der elektronen in kristallgittern. *Z. Phys.* **52**, 555–600 (1928).
4. Wildöer, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. Electronic structure of atomically resolved carbon nanotubes. *Nature* **391**, 59–62 (1998).
5. Bronikowski, M. J., Willis, P. A., Colbert, D. T., Smith, K. A. & Smalley, R. E. Gas-phase production of carbon single-walled nanotubes from carbon monoxide via the HiPco process: a parametric study. *J. Vac. Sci. Technol. A* **19**, 1800–1805 (2001).
6. Venema, L. C. *et al.* Length control of individual carbon nanotubes by nanostructuring with a scanning tunneling microscope. *Appl. Phys. Lett.* **71**, 2629–2631 (1997).
7. Tersoff, J. & Hamann, D. R. Theory of the scanning tunneling microscope. *Phys. Rev. B* **31**, 805–813 (1985).
8. Rubio, A., Sánchez-Portal, D., Artocho, E., Ordejón, P. & Soler, J. M. Electronic states in a finite carbon nanotube: a one-dimensional quantum box. *Phys. Rev. Lett.* **82**, 3520–3523 (1999).
9. Kane, C. L. & Mele, E. J. Broken symmetries in scanning tunneling images of carbon nanotubes. *Phys. Rev. B* **59**, R1279–R12762 (1999).
10. Venema, L. C. *et al.* Imaging electron wave functions of quantized energy levels in carbon nanotubes. *Science* **283**, 52–55 (1999).
11. Crommie, M. F., Lutz, C. P. & Eigler, D. M. Imaging standing waves in a two-dimensional electron gas. *Nature* **363**, 524–527 (1993).
12. Hasegawa, Y. & Avouris, Ph. Direct observation of standing wave formation at surface steps using scanning tunneling spectroscopy. *Phys. Rev. Lett.* **71**, 1071–1074 (1993).
13. Egger, R. & Gogolin, A. O. Effective low-energy theory for correlated carbon nanotubes. *Phys. Rev. Lett.* **79**, 5082–5085 (1997).
14. Kane, C., Balents, L. & Fisher, M. P. A. Coulomb interactions and mesoscopic effects in carbon nanotubes. *Phys. Rev. Lett.* **79**, 5086–5089 (1997).
15. Dresselhaus, G. *et al.* in *Science and Application of Nanotubes* (eds Tománek, D. & Endbody, R.) 275–295 (Kluwer Academic/Plenum, New York, 2000).
16. Venema, L. C. *et al.* Spatially resolved scanning tunneling spectroscopy on single-walled carbon nanotubes. *Phys. Rev. B* **62**, 5238–5244 (2000).
17. Liang, W. *et al.* Fabry-Perot interference in a nanotube electron waveguide. *Nature* **411**, 665–669 (2001).

Acknowledgements

We thank C. L. Kane for discussions and for making theoretical results known to us before publication. This work was supported by NASA, the Dutch Foundation for Fundamental Research (FOM) and the European Union IST-FET program SATURN.

Correspondence and requests for materials should be addressed to S.L. (e-mail: lemay@mb.tn.tudelft.nl).

Observation of individual vortices trapped along columnar defects in high-temperature superconductors

A. Tonomura^{*†}, H. Kasai^{*†}, O. Kamimura^{*†}, T. Matsuda^{*†}, K. Harada^{*†}, Y. Nakayama^{†‡}, J. Shimoyama^{†‡}, K. Kishio^{†‡}, T. Hanaguri^{†§}, K. Kitazawa^{†§}, M. Sasase^{||} & S. Okayasu^{||}

^{*} Advanced Research Laboratory, Hitachi Ltd, Hatoyama, Saitama 350-0395, Japan

[†] CREST, Japan Science and Technology Corporation (JST), Kawaguchi, Saitama 332-0012, Japan

[‡] Department of Applied Chemistry, University of Tokyo, Tokyo 113-8656, Japan

[§] Department of Advanced Materials Science, School of Frontier Sciences, University of Tokyo, Tokyo 113-0033, Japan

^{||} Department of Material Science, Japan Atomic Energy Research Institute, Tokai, Naka-gun, Ibaraki 319-1195, Japan

Many superconductors do not entirely expel magnetic flux—rather, magnetic flux can penetrate the superconducting state in the form of vortices. Moving vortices create resistance, so they must be ‘pinned’ to permit dissipationless current flow. This is a particularly important issue for the high-transition-temperature superconductors, in which the vortices move very easily¹. Irradiation of superconducting samples by heavy ions produces columnar defects, which are considered² to be the optimal pinning traps when the orientation of the column coincides with that of the vortex line. Although columnar defect pinning has been investigated using macroscopic techniques^{3,4}, it has hitherto been impossible to resolve individual vortices intersecting with individual defects. Here we achieve the resolution required to image vortex lines and columnar defects in Bi₂Sr₂CaCu₂O_{8+δ} (Bi-2212) thin films, using a 1-MV field-emission electron microscope⁵. For our thin films, we find that the vortex lines at higher temperatures are trapped and oriented along tilted columnar defects, irrespective of the orientation of the applied magnetic field. At lower temperatures, however, vortex penetration always takes place perpendicular to the film plane, suggesting that intrinsic ‘background’ pinning in the material now dominates.

There are several methods of directly observing vortices, but none of them can determine the behaviour of individual vortex lines inside superconductors. This is because these methods detect vortices at the superconductor surfaces, even though an attempt has been made to obtain evidence of wandering vortex lines near material defects by using a two-sided Bitter decoration technique to detect the vortex positions on both sides of the film⁶. At present, the only methods which enable the observation of individual vortices and defects inside superconducting thin films are Lorentz microscopy⁷ and interference microscopy⁸, where the vortex magnetic fields are detected using a penetrating electron beam. However, owing to the low penetration power of the existing 300-kV field-emission electron beam, only a film thinner than the diameter of the vortex magnetic flux (that is, twice the magnetic penetration depth) has been observed. Vortex lines oriented in different directions cannot be distinguished in such a thin film, because the vortex magnetic fields inside it do not change very much.

To obtain clear images of vortices inside thicker films, we have developed a large electron microscope⁵. This 40-ton microscope has a 1-MV field-emission electron beam that has more than twice the penetration power of a 300-kV beam and also the highest brightness ($2 \times 10^{10} \text{ A cm}^{-2} \text{ sr}^{-1}$) ever attained. These features of the electron beam have allowed us not only to observe vortices in Bi-2212 films 400 nm thick with high contrast, but also to

distinguish between two vortex lines⁹, one perpendicular to the film plane and the other trapped along a tilted columnar defect.

A schematic of the observation principle is shown in Fig. 1. Film samples were prepared by cleaving a single crystal of Bi-2212 (transition temperature $T_c = 85$ K) grown by the floating-zone method¹⁰. The films were then obliquely irradiated with parallel 240-MeV Au^{15+} ions (incidence angle $\theta_\phi = 70^\circ$) as sparsely as $0.05 \text{ ions } \mu\text{m}^{-2}$, or with a matching magnetic field of 0.1 mT. We intentionally made the column density extremely sparse in our experiments so that we could use untrapped vortex lines perpendicular to the film plane as reference lines to unambiguously distinguish vortex lines trapped along tilted columns. A collimated electron beam was applied incident to the tilted film (tilting angle $\alpha = 30^\circ$) and a magnetic field was applied in an arbitrary direction (incidence angle $\theta = (-70^\circ) - (+70^\circ)$) as shown in the schematic. We then observed the Lorentz images obtained by image-defocusing.

Examples of the observation results at $\theta_\phi = \theta = 70^\circ$ are shown in Fig. 2. In the in-focus image (Fig. 2a), the projected images of tilted columnar defects are seen as short black lines. As the defocusing distance Δf increases, the column images first disappear completely by image-blurring. At larger Δf values, spots with bright-and-dark contrasting features appear instead, as shown in Fig. 2b. These spots appear because the phase change of the transmitted electron beam due to the vortex magnetic flux is transformed into the visible intensity variations by image-defocusing. They are Lorentz images of vortices.

There are two kinds of images in this Lorentz micrograph: circular spots and elongated spots with lower contrast. The elongated spots are indicated by the arrows in Fig. 2b. Image simulation⁹ reveals that the circular images correspond to vortex lines penetrating the film perpendicularly to the film plane and that the elongated

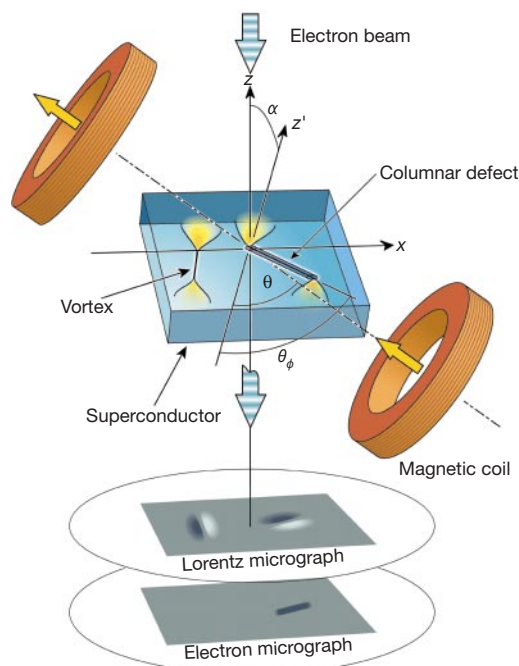


Figure 1 Lorentz microscopy of vortices in a superconducting Bi-2212 thin film with tilted columnar defects. The film is tilted at $\alpha = 30^\circ$, α being the rotation angle around the x axis between the optical axis (z axis) and the film normal (z' axis). A magnetic field is applied to the film in the x - z' plane at an incidence angle of θ , thus producing vortices, whereas that of the column direction is θ_ϕ . An electron beam is vertically applied incident to the film, and the phase change of the transmitted beam that is due to the vortex magnetic flux inside the film is transformed into intensity variations as a Lorentz image by image-defocusing. The vortex images appear to be different for perpendicular and tilted vortex lines under appropriate defocusing distances.

images correspond to vortex lines trapped along tilted columns. The image elongation direction is not necessarily the same as that of the tilted columns because of the effect of the inclination of the sample film, and the image contrast is weakened because of the spread of the magnetic flux distribution owing to the tilting of vortex lines from the c axis.

The vortex line direction inside a superconductor can be more precisely determined by comparing its Lorentz images taken while the sample is rotated around an axis normal to the film surface with the corresponding simulated images. A perpendicular vortex line does not change its direction with this rotation, but a tilted vortex line does, owing to the precession. In addition, a comparison of the two images in Fig. 2a and b reveals that columnar defects can be found at the exact centres of the elongated vortex images; this indicates that the elongated images actually represent vortex lines trapped along tilted columnar defects. Vortex lines oriented in different directions can thus be observed as different Lorentz images.

We then investigated whether vortex lines remained trapped along columnar defects even when the direction of the magnetic field or the sample temperature changed. First, we rotated the magnetic field from the column direction ($\theta = 70^\circ$) to the film normal direction ($\theta = 0^\circ$). In so doing, however, the elongated vortex images did not change at all, indicating that the vortex lines at columns remained trapped along them. Even when we further tilted the magnetic field direction until $\theta = -70^\circ$, the vortex lines remained trapped. Lorentz micrographs at $\theta = 70^\circ$ and $\theta = -70^\circ$ are shown in Fig. 3a and b. The elongated images are indicated by the arrows.

We also found that the elongated images remained unchanged when we increased T from 35 K to T_c . However, when we gradually decreased T , we found to our surprise that the trapped vortices at columnar defects began to “stand up” perpendicularly to the film plane between 12 and 14 K. Figure 3c shows a Lorentz micrograph of the vortices at $\theta = -70^\circ$ when T was 10 K. The elongated vortex images with weak contrast at 35 K (Fig. 3b) had changed to circular vortex images (Fig. 3c), the same as those surrounding untrapped perpendicular vortices.

We also found that this change was not reversible; there was hysteresis when the temperature was increased or decreased. That

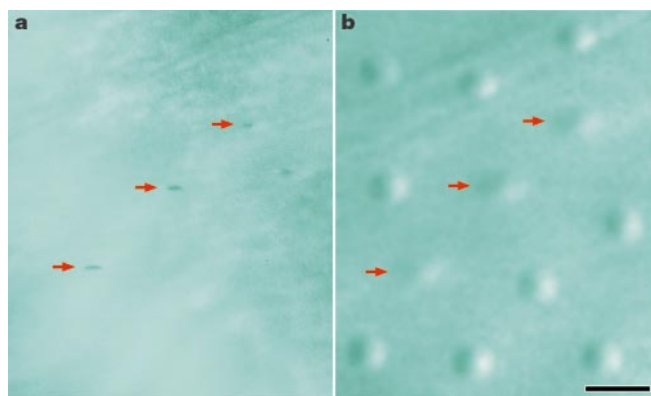


Figure 2 Vortices trapped and untrapped at columnar defects in Bi-2212 thin film ($H = 0.5$ mT, $\theta = \theta_\phi = 70^\circ$). **a**, Electron micrograph. **b**, Lorentz micrograph ($\Delta f = 500$ nm, $T = 30$ K). Columnar defects can be seen as short black lines in the in-focus electron micrograph (**a**). When the micrograph is greatly defocused, the column images completely disappear and the vortex images appear instead as can be seen in the Lorentz micrograph (**b**): the images of the perpendicular vortex lines are circular, whereas those of the tilted vortex lines are elongated and have lower contrast. This was confirmed by image simulation and is illustrated in Fig. 1. It should be noted here that the elongated vortex images (indicated by the arrows in (**b**)) appear at the positions of the column images in (**a**). Owing to the low density of columnar defects, only some vortex lines are trapped along columnar defects; the others are untrapped and penetrate the film perpendicularly to the film plane. Scale bar, 2 μm .

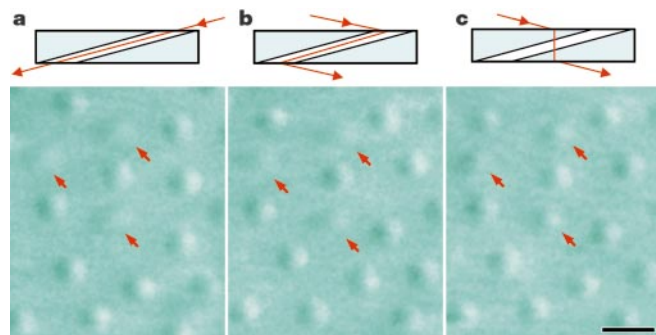


Figure 3 Lorentz micrographs showing vortex-line arrangements under various conditions. **a**, Magnetic field applied parallel to columns ($\theta = 70^\circ$) at $T = 35$ K. **b**, Magnetic field applied in the direction of $\theta = -70^\circ$ at $T = 35$ K. **c**, Magnetic field in the direction of $\theta = -70^\circ$ at $T = 10$ K. The vortex lines trapped along columnar defects (in **a**) remained trapped even when the direction of the magnetic field was greatly tilted, as can be seen in **b** (indicated by the arrow). When T decreased, however, the vortex lines trapped along tilted columnar defects began to penetrate the film perpendicularly to the film plane at 12–14 K, even though the vortex lines were located at the positions of the columns, and therefore the vortex images became circular and their contrast became higher, as can be seen in **c**. Illustrations above each Lorentz micrograph indicate vortex-line arrangements. Scale bar, 2 μm . Video clips showing the behaviour of the vortices are available at <http://www.hatoyama.hitachi.co.jp>.

is, when T was increased, perpendicular vortex lines at 7 K began to tilt along columnar defects at 15–19 K, rather than at 12–14 K. The transition temperatures were rather scattered for the different vortex lines. For example, even at 7 K, we found vortex lines tilted at columns in very rare cases. This does not seem to originate from the different pinning forces of individual columns, because electron microscopy showed that column structures were fairly uniform in shape. Such tilting at 7 K may be due to the fact that very strongly pinned vortices are trapped simultaneously at multiple columns that gather by chance rather than at a single isolated column.

The observed behaviour of vortices can be interpreted by taking into consideration the temperature dependence of the pinning forces of columnar defects estimated from the observation of vortex movements. Above 25 K, changes in the magnetic field caused sparsely distributed vortices to hop from one columnar defect to another when a driving force was applied to them. Such a movement occurred regardless of the direction in which the magnetic field was applied. This can happen when columnar-defect pinning is dominant. These results are consistent with the previous observation that the vortex lines are firmly trapped even along greatly tilted columns irrespective of the direction of the applied magnetic field at $T > 19$ K.

When T was decreased below 25 K, the hopping movement gradually changed in such a manner that vortices intermittently stopped and migrated. This kind of movement can be interpreted to occur, because the background collective pinning due to more abundant smaller atomic-size defects of other types (for example, oxygen defects) is thought to increase relative to the columnar-defect pinning. The migration speed decreases at lower T ; thus the migration is thought to be caused by thermally activated depinning of a vortex line from a large number of atomic-size defects one by one. The vortex lines, when stopped, were observed to be tilted along columns. When T was decreased below 12 K, however, the columnar-defect pinning became weaker relative to the background pinning, and trapped vortex lines at columnar defects were observed to stand up perpendicularly to the film plane. Around 7 K, all the vortices migrated or drifted uniformly when driven, as if the pinning force of columnar defects vanished and vortices were moving uniformly in a viscous medium. The pinning of columnar defects is almost completely hidden by the background pinning,

although there were still some exceptional vortices that were strongly pinned, possibly at multiple defects.

In this way, direct observation microscopically elucidated the behaviour of vortices in Bi-2212 thin films. The isotropic pinning of vortices by columnar defects at low magnetic fields indirectly inferred by macroscopic magnetization measurements^{3,11} can be explained microscopically by our experiments as follows: under our conditions of thin films, vortex lines are trapped along columns at temperatures above 19 K irrespective of the direction of the applied magnetic field; and at temperatures below 12 K vortex lines always penetrate the film perpendicularly to the film plane owing to the increased background pinning compared to the columnar-defect pinning and also because of the demagnetization effect of thin films. □

Received 27 March; accepted 4 June 2001.

1. Crabtree, G. W. & Nelson, D. R. Vortex physics in high-temperature superconductors. *Phys. Today* **50**, 38–45 (1997).
2. Civale, L. *et al.* Vortex confinement by columnar defects in $\text{YBa}_2\text{Cu}_3\text{O}_7$ crystals: Enhanced pinning at high fields and temperatures. *Phys. Rev. Lett.* **67**, 648–651 (1991).
3. Hardy, V. *et al.* Accommodation of vortices to tilted line defects in very high- T_c superconductors with various electronic anisotropies. *Phys. Rev. B* **54**, 656–664 (1996).
4. Schuster, Th. *et al.* Observation of in-plane anisotropy of vortex pinning by inclined columnar defects. *Phys. Rev. B* **50**, 9499–9502 (1994).
5. Kawasaki, T. *et al.* Fine crystal lattice fringes observed using a transmission electron microscope with 1 MeV coherent electron waves. *Appl. Phys. Lett.* **76**, 1342–1344 (2000).
6. Yao, Z., Yoon, S., Dai, H., Fan, S. & Lieber, C. M. Path of magnetic flux lines through high- T_c copper oxide superconductors. *Nature* **371**, 777–779 (1994).
7. Harada, K. *et al.* Real-time observation of vortex lattices in a superconductor by electron microscopy. *Nature* **360**, 51–53 (1992).
8. Tonomura, A. *Electron Holography* 2nd edn (Springer, Heidelberg, 1999).
9. Fanesi, S. *et al.* Influence of core misalignment and distortion on the Fresnel and holographic images of superconducting fluxons. *Phys. Rev. B* **59**, 1426–1431 (1999).
10. Kotaka, Y. *et al.* Doping state and transport anisotropy in Bi2212 single crystals. *Physica C* **235–240**, 1529–1530 (1994).
11. Klein, L., Yacoby, E. R., Yeshurun, Y., Konczykowski, M. & Kishio, K. Evidence for line vortices in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$. *Phys. Rev. B* **48**, 3523–3525 (1993).

Acknowledgements

We are grateful to N. Hatano for his discussions on tilted columnar defects, and to G. Pozzi, M. Beleggia, N. Osakabe, T. Yoshida and J. Masuko for their simulations of Lorentz micrographs of tilted vortex lines. We also thank F. Nori for discussions and T. Akashi and I. Matsui for help with the experiments.

Correspondence and requests for materials should be addressed to A.T. (e-mail: tonomura@harl.hitachi.co.jp).

Vacancies in solids and the stability of surface morphology

K. F. McCarty, J. A. Nobel & N. C. Bartelt

Sandia National Laboratories, Livermore, California 94551-0969, USA

Determining how thermal vacancies are created and destroyed in solids is crucial for understanding many of their physical properties, such as solid-state diffusion. Surfaces are known to be good sources and sinks for bulk vacancies, but directly determining where the exchange between the surface and the bulk occurs is difficult. Here we show that vacancy generation (and annihilation) on the (110) surface of an ordered nickel–aluminium intermetallic alloy does not occur over the entire surface, but only near atomic step edges. This has been determined by oscillating the sample's temperature and observing in real time the response of the surface structure as a function of frequency (a version of Ångström's method of measuring thermal conductivity¹) using low-energy electron microscopy. Although

the surface-exchange process is slow compared with bulk diffusion, the vacancy-generation rate nevertheless controls the dynamics of the alloy surface morphology. These observations, demonstrating that surface smoothing can occur through bulk vacancy transport rather than surface diffusion, should have important implications for the stability of fabricated nanoscale structures.

As it is extremely difficult to prepare perfectly flat crystalline surfaces, solid surfaces usually contain arrays of atomic steps. These step arrays are not in equilibrium because a surface can lower its free energy by decreasing the total step length. Figure 1a illustrates this lack of stability on a NiAl (110) surface as observed at 957 °C with low-energy electron microscopy (LEEM). The dark lines are steps of a single atomic layer in height, arranged in a concentric array of islands, similar to a wedding cake's tiers. The islands in the stack shrink (Fig. 1b) with time, reducing the surface's step length. Thus, the surface lowers its free energy and, eventually, becomes smoother.

Although surface smoothing has been studied extensively over the past 50 years^{2,3}, only recently have experimental probes such as LEEM⁴ been able to observe atomic-step motion in real time, as in Fig. 1. To understand the smoothing mechanism, characterizing the diffusional atomic currents originating from shrinking step edges is essential. In this regard, Fig. 1 has a remarkable feature: all steps in

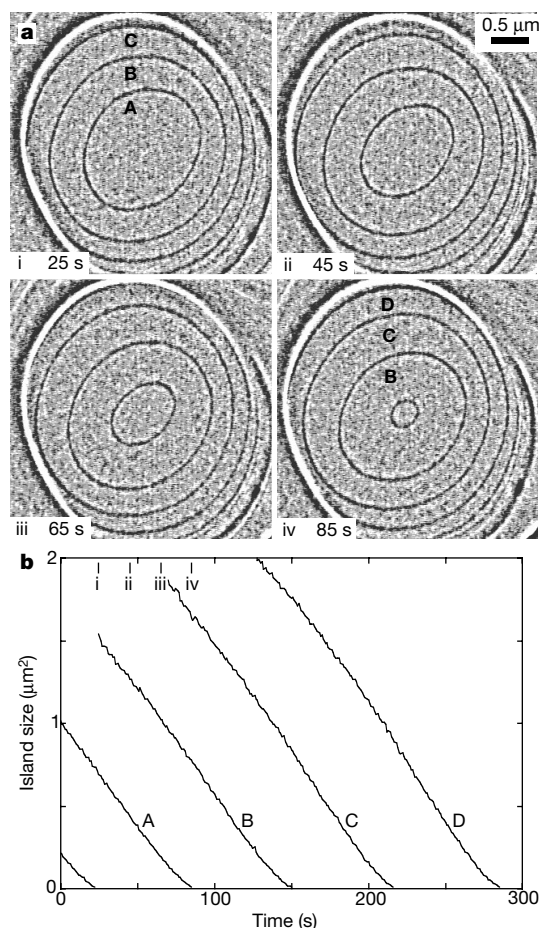


Figure 1 Smoothing of the NiAl (110) surface. **a**, Low-energy electron micrographs captured during the constant-temperature decay of an island-stack structure on the NiAl (110) surface at 957 °C. The dark lines are the monoatomic surface steps. All images obtained in bright-field mode using 3.8 eV electrons. (An example of island decay is given in Movie 1 of the Supplementary Information.) **b**, Sizes (square of the largest dimension) of the islands labelled in part **a** as a function of time. The times at which the images of **a** were captured are marked. Despite the large changes in the local step environment and the different curvatures of the islands, the islands all decay at the same, constant rate.

the field of view are shrinking and it is not clear where the emitted atoms are going. If surface smoothing occurs by surface diffusion, as usually assumed, the atomic currents are between neighbouring step edges⁵. The surface then lowers its free energy by shrinking small islands at the expense of growing large islands⁶. But there is no evidence for these currents on NiAl (110), as all islands shrink at the same rate (Fig. 1b) at this temperature, regardless of each island's environment. For example, when the stack's topmost island and thus the emitted current vanishes, the decay rate of the remaining islands is unchanged.

To explain the surface-smoothing mechanism, we show that the currents responsible for smoothing are vacancies being transported into and out of the solid. To prove this, we first establish that there is a ready exchange of bulk vacancies at step edges. Figure 2a shows LEEM images of an island stack as the temperature is oscillated. With a temperature increase or decrease, respectively, the topmost island markedly grows (Fig. 2a i–ii) or shrinks (Fig. 2a ii–iii). Complementary behaviour is observed for vacancy islands (that is, single-atomic-layer pits): they shrink or enlarge with a temperature

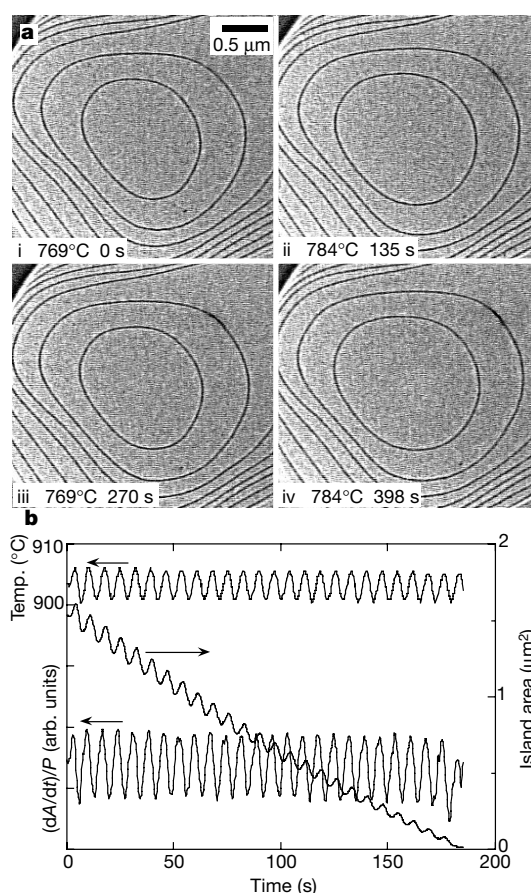


Figure 2 Measurement of bulk-vacancy fluxes towards atomic step edges. **a**, Low-energy electron micrographs of an island-stack structure on the NiAl (110) surface as the temperature is varied sinusoidally about a mean of 776 °C. With increasing temperature, bulk atoms come to the surface because the concentration of thermal vacancies in the bulk increases. For decreasing temperature, the islands shrink in size because material leaves the surface to reduce the bulk vacancy concentration. (An additional example is given in Movie 2 of the Supplementary Information.) **b**, Island area $A(t)$ and the rate of change normalized by the perimeter $(dA/dt)/P$ as the temperature is sinusoidally oscillated by about ± 2.5 °C around a mean of 903 °C. Even though the island area decreases greatly because of thermal smoothing, the area change resulting from the temperature oscillation scales exactly as the perimeter of the island and is independent of the size of this island and the surrounding concentric islands. The surface dynamics, therefore, are dominated by processes occurring near the steps and are independent of the terrace spacings.

increase or decrease, respectively. This phenomenon is even more marked for larger temperature changes: the Supplementary Information shows examples of at least seven layers being added to or removed from the surface. Because a surface can provide only a fraction of a layer, explanations based solely on surface processes are ruled out. Instead, we attribute these size variations to the fact that the equilibrium concentration of vacancies in the bulk changes with temperature. To restore equilibrium in the bulk, vacancies are either created or annihilated at the surface, causing the surface steps to move. The resulting concentration gradient perpendicular to the surface produces a net mass flux out of the bulk (during heating) or into the bulk (during cooling) that attempts to equilibrate the crystal through vacancy creation or annihilation at the surface.

We next show that vacancies are generated and annihilated predominantly near step edges. We do this by proving that the vacancy flux to each step is simply proportional to step length. Figure 2b depicts an island's size as the temperature is oscillated sinusoidally. The figure also shows that the time derivative of the island area (dA/dt) divided by the island perimeter is a sinusoidally varying function of constant amplitude even though the island area decreases greatly from thermal smoothing. (The time derivative separates the oscillating component from the slow decay due to thermal smoothing.) The remarkably simple conclusion that islands capture and lose atoms simply in proportion to their step length holds for all islands examined at all observed temperatures (about 740–940 °C) independent of the island's environment (such as the density of nearby steps). This establishes that vacancies are generated only near steps. Although steps are expected to be the ultimate source or sink for the bulk vacancies (as has been suggested for surface vacancies⁷), that the steps would control the rate at which the bulk vacancies are created or annihilated is not obvious. For example, the vacancies could have been created everywhere on the

terraces with the emitted atoms rapidly diffusing to the steps.

By measuring the frequency dependence of the amplitude of the area oscillations at different temperatures, we can estimate the vacancy formation and migration energies. Motivated by the fact that the decay rates in Fig. 1 lack an environmental dependence, we first make the assumption (justified below) that the only significant gradients in vacancy concentration are perpendicular to the surface. We further assume that the only source and sinks of vacancies are at the surface. We then solve the one-dimensional diffusion equation $\partial c_B/\partial t = D\partial^2 c_B/\partial x^2$ for a material slab of width L subject to the boundary conditions that the time-dependent vacancy concentrations $c_B(x, t)$ at both surfaces are oscillating with frequency ω : $c_B(\pm L/2, t) = c_0 \cos(\omega t)$. (This boundary condition appropriately describes the large-scale currents into the bulk of the solid that occur at timescales for which the surface region, but not the bulk, is already close to equilibrium.) The flux to the surface F is then found to have a phase shift δ with respect to the temperature oscillations:

$$F = -D \left(\frac{\partial c_B}{\partial x} \right)_{x=L/2} = F_0 \cos(\omega t - \delta)$$

where D is the diffusion coefficient and δ is

$$\delta(k, L) = \arctan \left[\frac{\sin(kL) + \sinh(kL)}{\sin(kL) - \sinh(kL)} \right] \quad (1)$$

where $k = \sqrt{\omega/2D}$. The normalized flux amplitude F_0/ω is:

$$\frac{F_0(k, L)}{\omega} = \frac{c_0}{\sqrt{2}k} \frac{1 [\sinh^2(kL) + \sin^2(kL)]^{1/2}}{\cos(kL) + \cosh(kL)} \quad (2)$$

The change in area $\Delta A(t)$ of an island with perimeter P is proportional to the product of the time integral of the surface flux and P :

$$\Delta A(t) \propto P \frac{F_0}{\omega} \cos(\omega t - \delta - \pi/2) \quad (3)$$

In the high and low frequency limits, $\delta \rightarrow -\pi/4$ and $\delta \rightarrow -\pi/2$, respectively. Figure 3a shows that the one-dimensional model is appropriate: the observed phase shift (equation (1)) and the change in island size F_0 (equation (2)) are both well described using the same fitted value of $L^2/2D = 149$ s. Taking $L = 1.5$ nm, the crystal thickness, gives D an upper bound of $7.5 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$. (This is only an upper limit because there may be vacancy sources other than the surface.) By repeating this analysis at other temperatures, we have determined the activation energy of D to be about 1 eV.

We determine the vacancy-formation energy (E_F) by measuring the flux as a function of temperature. The thermal bulk vacancy concentration is expected to be proportional to $\exp(-E_F/kT)$. The concentration change Δc_B produced by a temperature change ΔT is then proportional to $\Delta T \exp(-E_F/kT)/T^2$. As an island's area change is proportional to Δc_B and (from equation (3)) is also proportional to F_0/ω , we have, in the limit of small frequencies,

$$\frac{\Delta c_B T^2}{\Delta T} = \frac{F_0 T^2}{\omega \Delta T} \propto e^{-E_F/kT}$$

which allows E_F to be determined from measurements of F_0 .

We consider data for F_0 with frequencies small enough that $\delta \leq -0.42\pi$. (The diffusion model implies that these oscillation rates are slow enough to ensure that >95% of the vacancies that are necessary to equilibrate fully the crystal are produced at the surface.) Plotting $\ln(F_0 T^2/\omega \Delta T)$ against $1/T$ gives a line with $E_F = 0.64$ eV (Fig. 3b). Because our crystal is nickel-rich ($\text{Ni}_{57}\text{Al}_{43}$), the dominant zero-temperature defects are Ni atoms on the Al sublattice of the caesium chloride structure⁸. The dominant thermal defect is the 'triple defect'⁹ (two Ni vacancies and a Ni atom on an Al site), whose formation moves a bulk NiAl unit to the surface^{10–14}. (Interstitials are not important in intermetallic alloys¹⁰.) Because the NiAl (110) surface has the bulk composition¹⁵ and we see no image

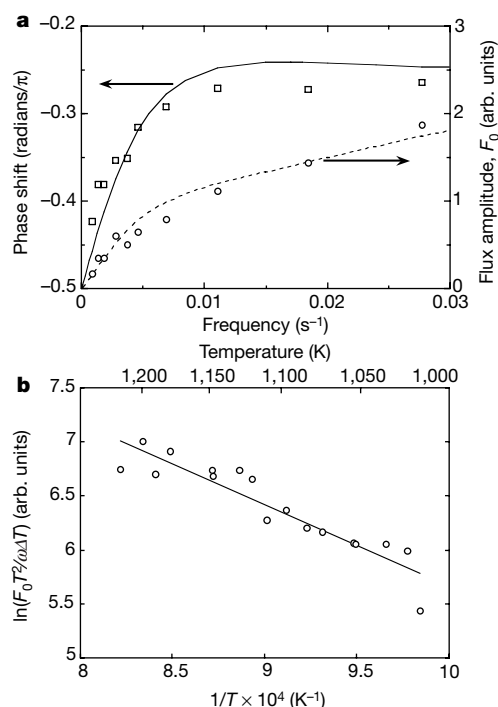


Figure 3 Interpretation of the frequency dependence of the vacancy fluxes to the surface. **a**, Phase shift and amplitude of island areas as a function of frequency for temperature oscillations around 777 °C. The symbols are the experimental data and the lines result from using a single diffusion constant to fit both expressions from the one-dimensional bulk-diffusion model. **b**, Arrhenius plot of the normalized flux that comes from the bulk to the surface for a given temperature change. The slope gives the effective defect formation energy, which is 0.64 eV.

contrast associated with changing stoichiometry, the material being added to the surface is NiAl alloy. Thus, thermal defects leave the bulk composition unchanged. In this case, E_F is exactly one-half the energy required to make the triple defect^{10–14} and thus is an ‘effective’ defect-formation energy. (Our triple-defect formation energy (1.28 eV) is somewhat lower than other less-direct experimental values for Ni-rich NiAl (1.65–1.83 eV; ref. 16) and considerably lower than that from theoretical calculations (2.0–2.4 eV; refs 11–14). Given the apparent simplicity of our approach, this discrepancy with theory is puzzling.) The vacancy concentration in Al-rich crystals has been predicted to decrease with increasing temperature, rather than increase^{11,17}. As this behaviour would give area oscillations that are 180° out-of-phase with those observed for our Ni-rich crystal, our technique can be used to verify such predictions.

The one-dimensional diffusion model’s success in simulating the phase shift and amplitude and measuring the vacancy formation energy provides further strong evidence that the only large gradients in vacancy concentrations are perpendicular to the surface and that lateral gradients are small. This lack of concentration gradients parallel to the surface implies that step-vacancy exchange is slow compared with bulk diffusion: any (nanoscale) gradients parallel to the surface caused by vacancy emission from step edges must be quickly eliminated by fast vacancy diffusion. As a consequence, adjacent surface features of different curvature cannot transfer mass through a near-surface current because there are no lateral gradients, consistent with the decay rates being independent of environment (Fig. 1).

The functional form of the island decay rates (Fig. 1b) also supports the conclusion that lateral concentration gradients are negligible: because each step edge interacts with the same uniform reservoir of vacancies, smoothing should then follow the same kinetics as if steps were exchanging atoms with a uniform vapour phase¹⁸. We assume that the rate of random-exchange vacancy creation and annihilation near a straight step edge is $h_{B-S}c_B$ (that is, proportional to the vacancy concentration). Using the Gibbs–Thomson chemical potential of a curved step, one then finds that a circular island of radius R will have a constant rate of area change:

$$\frac{dA}{dt} = - \frac{2\pi a^4 h_{B-S} c_B \beta}{kT} = \text{constant}$$

where a^2 is atomic size, and β is step free energy per length. We find that islands do indeed shrink at a constant rate when they are small enough (Fig. 1b). (Vacancy islands also fill in at the same constant rates as normal islands.) We emphasize that the observed kinetics suggest that the isothermal smoothing of the NiAl (110) surface is not controlled by the rate of bulk transport. Instead, smoothing is controlled by the rate at which bulk vacancies exchange near steps. The rates of island decay show Arrhenius behaviour with an activation energy of 2.54 eV. As the vacancy formation energy is 0.64 eV, the rate of vacancy exchange between bulk and surface steps (h_{B-S}) has an activation energy of 1.9 eV. The exchange near surface steps is slow compared with bulk diffusion because this energy is considerably higher than the energies required to thermally generate (0.64 eV) and move (about 1 eV) vacancies in the bulk. (During isothermal surface smoothing, only the surface is out of equilibrium. In contrast, after a temperature change, the entire sample is not equilibrated. How fast the bulk then equilibrates is controlled by the rate at which vacancies diffuse through the bulk, as described by the diffusion model above.)

Although the idea that surface dynamics are dominated by diffusion through the bulk was proposed early on by Herring² and Mullins³, most current work assumes that the important processes occur exclusively through surface diffusion in the first one or two layers⁵. The lack of environmental dependence of island decay rates argues against surface diffusion playing a role on NiAl (110), even at about half its melting point (1,640 °C; ref. 8). (In principle,

extremely ‘permeable’¹⁹ steps could cause island decay rates to be environment-independent. However, our observation of vacancy-step exchange provides the more plausible explanation here.) We expect that similar behaviour will be directly observed on other alloys and elemental metals. Because bulk vacancies have larger formation and diffusion activation energies than surface vacancies, bulk processes should become more important at high temperature³. To understand this, consider the root-mean-square displacement of a surface vacancy or adatom before it becomes or annihilates a bulk vacancy. When this length, which will decrease with increasing temperature, becomes much smaller than the feature size, bulk diffusion will dominate. In fact, early experimental work analysing decay of periodic surface profiles suggested that bulk diffusion contributed significantly to surface flattening of Au (ref. 20), Ni (ref. 21), and Cu (ref. 22) near their melting points. That vacancies on the NiAl (110) surface are generated and annihilated only near steps undoubtedly affects how other processes involving vacancies occur (for instance, surface oxidation). Finally, imaging the frequency-dependent response of nanoscale surface structure to temperature variations is a promising technique for measuring both surface and bulk diffusion constants. □

Received 19 February; accepted 15 June 2001.

1. Ångström, A. J. New method of determining the thermal conductivity of bodies. *Phil. Mag.* **25**, 130–143 (1863).
2. Herring, C. in *Structure and Properties of Solid Surfaces* (eds Gomer, R. & Smith, C. S.) (Univ. Chicago Press, 1952).
3. Mullins, W. W. Flattening of a nearly plane solid surface due to capillarity. *J. Appl. Phys.* **30**, 77–83 (1959).
4. Bauer, E. Low-energy-electron microscopy. *Rep. Prog. Phys.* **57**, 895–938 (1994).
5. Jeong, H. C. & Williams, E. D. Steps on surfaces: experiment and theory. *Surf. Sci. Rep.* **34**, 175–294 (1999).
6. Giesen, M. & Ibach, H. Step edge barrier controlled decay of multilayer islands on Cu(111). *Surf. Sci.* **431**, 109–115 (1999).
7. Ibach, H., Giesen, M., Flores, T., Wuttig, M. & Treglia, G. Vacancy generation at steps and the kinetics of surface alloy formation. *Surf. Sci.* **364**, 453–466 (1996).
8. Bradley, A. J. & Taylor, A. An X-ray analysis of the nickel–aluminium system. *Proc. R. Soc. Lond. A* **159**, 56–72 (1937).
9. Wasilewski, R. J. Structure defects in CsCl intermetallic compounds—I theory. *J. Phys. Chem. Solids* **29**, 39–49 (1968).
10. Hagen, M. & Finnis, M. W. Point defects and chemical potentials in ordered alloys. *Phil. Mag. A* **77**, 447–464 (1998).
11. Korzhavyi, P. A. et al. Constitutional and thermal point defects in B2 NiAl. *Phys. Rev. B* **61**, 6003–6018 (2000).
12. Meyer, B. & Fahnle, M. Atomic defects in the ordered compound B2-NiAl: A combination of *ab initio* electron theory and statistical mechanics. *Phys. Rev. B* **59**, 6072–6082 (1999).
13. Meyer, B. & Fahnle, M. Atomic defects in the ordered compound B2-NiAl: A combination of *ab initio* electron theory and statistical mechanics [(*Phys. Rev. B* **59**, 6072 (1999)). *Phys. Rev. B* **60**, 717–717 (1999)].
14. Mishin, Y. & Farkas, D. Atomistic simulation of point defects and diffusion in B2 NiAl. I. Point defect energetics. *Phil. Mag. A* **75**, 169–185 (1997).
15. Davis, H. L. & Noonan, J. R. Rippled relaxation in the (110) surface of the ordered metallic alloy NiAl. *Phys. Rev. Lett.* **54**, 566–569 (1985).
16. Bai, B. & Collins, G. S. in *High-Temperature Ordered Intermetallic Alloys VII, MRS Symposia Proceedings No. 552* (eds George, E. P., Yamaguchi, M. & Mills, M. J.) KK8.7.1 (Materials Research Society, Pittsburgh, 1999).
17. Smirnov, A. A. Anomaly in the temperature dependence of the vacancy concentration in NiAl type alloys. *Sov. Phys. Dokl.* **36**, 479–480 (1991).
18. Zinke-Allmang, M., Feldman, L. C. & Grabow, M. H. Clustering on surfaces. *Surf. Sci. Rep.* **16**, 377–463 (1992).
19. Tanaka, S., Bartelt, N. C., Umbach, C. C., Tromp, R. M. & Blakely, J. M. Step permeability and the relaxation of biperiodic gratings on Si(001). *Phys. Rev. Lett.* **78**, 3342–3345 (1997).
20. Blakely, J. M. Effect of impurity on surfaces of heated gold. *Trans. Faraday Soc.* **57**, 1164–1168 (1961).
21. Maiya, P. S. & Blakely, J. M. Surface self-diffusion and surface energy of nickel. *J. Appl. Phys.* **38**, 698–704 (1966).
22. Hoehne, K. & Sizmann, R. Volume and surface self-diffusion measurements on copper by thermal surface smoothing. *Phys. Status Solidi A* **5**, 577–589 (1971).

Supplementary information is available on Nature’s World-Wide Web site (<http://www.nature.com>).

Acknowledgements

This work was supported by the Office of Basic Energy Sciences, Division of Materials Sciences of the US Department of Energy. We thank D. C. Dibble, N. Y. C. Yang and K. J. Gross for technical assistance and B. Poelsema, J. C. Hamilton, J. J. Hoyt, J. B. Hannon, P. J. Feibelman and G. E. Thayer for discussions.

Correspondence should be addressed to K.F.M. (e-mail: mccarty@sandia.gov).

Generation and characterization of a fairly stable triplet carbene

Hideo Tomioka, Eri Iwamoto, Hidetaka Itakura & Katsuyuki Hirai

Chemistry Department for Materials, Faculty of Engineering, Mie University, Tsu, Mie 514-8507, Japan

Most molecules are held together by covalent bonds—electron pairs jointly shared by the two atoms that are linked by the bond. Free radicals, in contrast, have at least one unpaired electron. In the case of carbon-based radicals, the carbon atom at the radical centre no longer makes four bonds with other atoms as it would do in its normal, tetravalent state. The presence of unpaired electrons renders such radicals highly reactive, so they normally occur only as transient intermediates during chemical reactions. But the discovery^{1,2} by Gomberg in 1900 of triphenylmethyl, the first relatively stable free radical containing a central trivalent carbon atom, illustrated that radicals with suitable geometrical and electronic structures can be stable. Compounds containing a divalent carbon atom that uses only two of its four valence electrons for bonding are usually less stable than Gomberg-type radicals with trivalent carbon^{3–5}. Although the role of these so-called carbenes in chemical reactions has long been postulated, they were unambiguously identified only in the 1950s. More recently, stable carbenes have been prepared^{6,7}, but the singlet state of these molecules^{6–12}, with the two nonbonding valence electrons paired, means that they are not radicals. Carbenes in the second possible electronic state, the triplet state, are radicals: the two nonbonding electrons have parallel spins and occupy different orbitals^{13,14}. Here we report the preparation and characterization of a triplet carbene, whose half-life of 19 minutes at room temperature shows it to be significantly more stable than previously observed triplet carbenes^{15–17}.

Triplet carbenes are usually characterized by electron paramagnetic resonance (EPR) spectroscopy. The principal information extracted from the EPR spectra of triplet carbenes are the zero-field splitting parameters, D and E (interaction energies in cm^{-1}), which measure the magnetic dipole interaction of the unpaired electrons in the absence of an external field. Usually D provides a measure of the average distance r between the unpaired electrons; in carbenes with conjugated π -systems, it thus allows for a qualitative

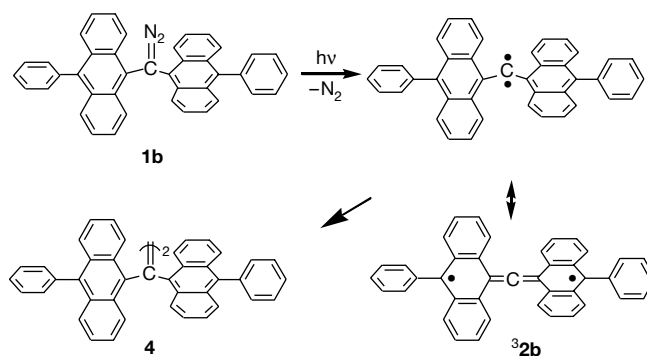


Figure 2 Structure of triplet bis(9-anthryl)carbene (**³2b), generated by photolysis of the precursor diazomethane (**1b**), and the corresponding carbene dimer (**4**). The phenyl groups in **³2b** are not expected to be in the same plane as the anthryl rings owing to the repulsion between *ortho*- and *peri*-hydrogens. This is in line with the similarity of the zero-field splitting parameters of **³2b** to those of **³2a**. It seems that the phenyl groups effectively prevent the unpaired electrons from 'leaking out' and thus prevent reaction at position 10.**

determination of the extent of electron delocalization. E provides a measure of the difference in the magnitude of magnetic dipole moment among triplet species, from which the bond angle at the carbene centre can be estimated. In the case of diarylcarbenes, E and D values are expected to decrease as the carbene bond angle expands and the unpaired electrons are more extensively delocalized^{18–21}.

Among diarylcarbenes, triplet di(9-anthryl)carbene (**³2a) exhibits the smallest D (0.113 cm^{-1}) and E (0.0011 cm^{-1}) values reported²², indicating that it has an almost linear bond geometry at the carbene centre with extensive delocalization of the unpaired electrons onto the perpendicularly attached anthryl groups. The extensive delocalization is expected to stabilize this carbene compound thermodynamically, while the perpendicular geometry of the anthryl groups stabilizes the carbene centre kinetically, through shielding with the four *peri*-hydrogens. That is, the electronic factor and the molecular structure of the molecule seem ideal for the formation of a stable triplet carbene. In spite of those highly favourable structural factors, **³2a is very ephemeral²³: its lifetime in degassed benzene solution at room temperature is $0.5 \mu\text{s}$, shorter even than that of triplet diphenylcarbene.****

Product analysis shows that **³2a forms as main product a trimer (**3**) (ref. 24), in which three units of **³2a are connected at the C_{10}****

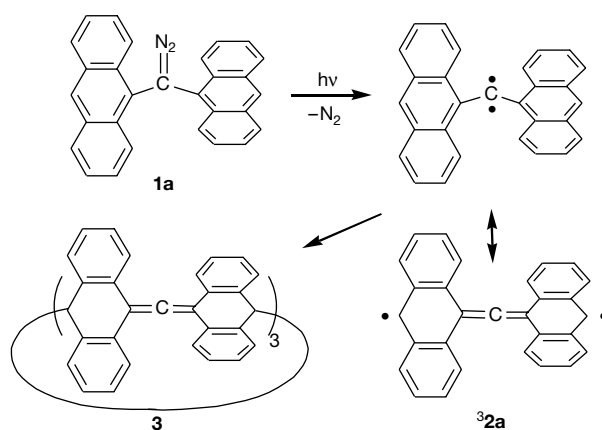


Figure 1 Structure of triplet di(9-anthryl)carbene (**³2a), generated by photolysis of the precursor diazomethane (**1a**), and the corresponding trimer (**3**). In **³2a**, the carbene centre is almost linear, and there is extensive delocalization of the unpaired electrons onto the anthryl groups, which are attached in a perpendicular fashion. These electronic and**

structural factors would be expected to favour the formation of a stable triplet carbene. However, **³2a is very ephemeral. The main decay pathway of **³2a is reaction at the C_{10} position of the anthryl groups in a cyclic way to form a trimer (**3**).****

position of the anthryl groups in a cyclic way. This behaviour suggests that delocalization of the unpaired electrons in $^3\mathbf{2a}$ leads to their 'leaking out' from the carbene centre to position 10, where sufficient spin density builds up for the trimerization to take place (Fig. 1). At the same time, the lack of formation of olefin-type dimers through coupling two units of $^3\mathbf{2a}$ at their carbene centres indicates that the carbene centre itself is indeed well shielded and stabilized.

The above observations indicate that the stability of $^3\mathbf{2a}$ would increase if the trimerization reaction were somehow suppressed—by, for example, the introduction of a substituent at position C₁₀ to block the reactivity there. This simple idea inspired us to synthesize $^3\mathbf{2b}$, which is formally related to $^3\mathbf{2a}$ by substituting a phenyl group at each of the two C₁₀ positions.

Irradiation (at wavelengths $\lambda > 300$ nm) of bis[9-(10-phenyl)anthryl]diazomethane $\mathbf{1b}$ in a 2-methyltetrahydrofuran (2-MTHF) glass at 77 K gave EPR signals very similar to those observed for dianthrylcarbene $^3\mathbf{2a}$ (Fig. 2). The zero-field splitting parameters ($D = 0.105$ cm⁻¹, $E = 4.4 \times 10^{-4}$ cm⁻¹) derived from the signals are essentially the same as those obtained for the triplet dianthrylcarbene, indicating that the phenyl groups are not in the same plane as the anthryl rings owing to the repulsion between *ortho*- and *peri*-hydrogens. The signals of $^3\mathbf{2a}$ and $^3\mathbf{2b}$ differ only in their thermal response: when the 2-MTHF glass containing $^3\mathbf{2}$ was warmed gradually, the signals due to $^3\mathbf{2a}$ started to disappear at around 90 K, whereas no significant decay of the signals of $^3\mathbf{2b}$ was observed up to 240 K. The signal of $^3\mathbf{2b}$ started to decay only at around 270 K ($\sim 0^\circ\text{C}$), but was still visible when heating to 300 K ($\sim 27^\circ\text{C}$) (see Supplementary Information).

The EPR signals of $^3\mathbf{2b}$ became sharp and shifted slightly but distinctly at around 110 K. Changes of this kind have often been observed for sterically congested triplet diarylcarbenes; they are usually attributed to the relaxation of the carbene from distorted orientations—caused by the trapping of the precursor in the frozen matrix—to the thermodynamically favourable orientation upon softening of the matrix²⁵.

Ultraviolet–visible (UV/vis.) spectroscopic studies provided similar but more quantitative results^{21,26–28}. Irradiation ($\lambda > 300$ nm) of $\mathbf{1b}$ in the 2-MTHF matrix at 77 K resulted in the appearance of new absorption bands (343, 362 and 475 nm). As under identical conditions strong EPR signals ascribable to $^3\mathbf{2b}$ were observed, the UV/vis absorption bands were assigned to $^3\mathbf{2b}$. When the temperature of the matrix containing $^3\mathbf{2b}$ was raised gradually, a

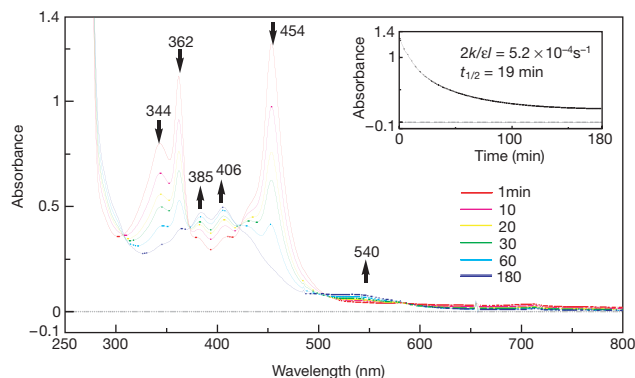


Figure 3 Ultraviolet/visible spectrum after photolysis ($\lambda = 308$ nm) of the precursor diazomethane ($\mathbf{1b}$) in degassed benzene at 20°C . Triplet bis[9-(10-phenyl)anthryl]carbene ($^3\mathbf{2b}$) showed transient absorption bands at 344, 362 and 454 nm. The spectra recorded after 1, 10, 20, 30, 60 and 180 min are shown. Inset, the decay of the transient absorption due to $^3\mathbf{2b}$ monitored at 454 nm. The bands due to $^3\mathbf{2b}$ decayed cleanly, showing isosbestic points, and persisted for more than 3 hours before disappearing completely. k , rate constant (in $\text{mol}^{-1} \text{dm}^3 \text{s}^{-1}$); ϵ , molar extinction coefficient of $^3\mathbf{2b}$ (in $\text{mol}^{-1} \text{dm}^3 \text{cm}^{-1}$); l , cell-path length (in cm); $t_{1/2}$, half-life (in min).

broad absorption maximum at 475 nm became sharp and shifted to 454 nm at 110 K. As this is the temperature where the EPR signals shifted due to geometrical relaxation of $^3\mathbf{2b}$, we assigned these new bands to the relaxed $^3\mathbf{2b}$. When the temperature was further raised, the absorption bands decayed very slowly and cleanly, showing isosbestic points. These were observed even at 270 K, and did not disappear completely even at 300 K (see Supplementary Information). Product analysis of the spent solution showed the presence of a carbene dimer ($\mathbf{4}$) as the main isolable product (Fig. 2).

To probe the stability and decay kinetics of $^3\mathbf{2b}$ at ambient conditions, $\mathbf{1b}$ was photolysed in degassed benzene at room temperature. The transient absorption bands observed were identical to those observed for the relaxed $^3\mathbf{2b}$ in the low-temperature matrix. The bands due to $^3\mathbf{2b}$ decayed cleanly, showing isosbestic points; they persisted for more than 3 hours before disappearing completely. Fitting the decay curve with second-order kinetics ($2k/\epsilon l = 5.2 \times 10^{-4} \text{s}^{-1}$, where k is the rate constant (in $\text{mol}^{-1} \text{dm}^3 \text{s}^{-1}$), ϵ is the molar extinction coefficient of $^3\mathbf{2b}$ (in $\text{mol}^{-1} \text{dm}^3 \text{cm}^{-1}$) and l is the cell-path length (in cm), a half-life for $^3\mathbf{2b}$ of 19 min was estimated (Fig. 3).

One of the resonance structures of $^3\mathbf{2b}$ can formally be regarded as two Gomberg radicals connected by an allene bond (see Fig. 4). Small D and E values and an exclusive coupling reaction at the 10 positions (in the absence of the phenyl substituents) suggest that the contribution of this structure to the actual average distribution of electrons in the molecule is important, raising the question of whether $^3\mathbf{2}$ is a triplet carbene or a triplet diradical. In general, D values decrease as the unpaired electrons are delocalized. However, the information contained in the D values is insufficient to discriminate between a carbene (1,1-diradical) and a diradical (containing unpaired electrons at two discrete centres). However, we note that the measured E values indicate that $^3\mathbf{2b}$, and thus also its central bond angle, is not completely linear; $^3\mathbf{2b}$ must therefore have some carbene characteristics.

Because spectroscopic methods seem unable to discriminate between a carbene and a diradical species, we use chemical reactivity to probe whether the unpaired electrons are localized on one carbon

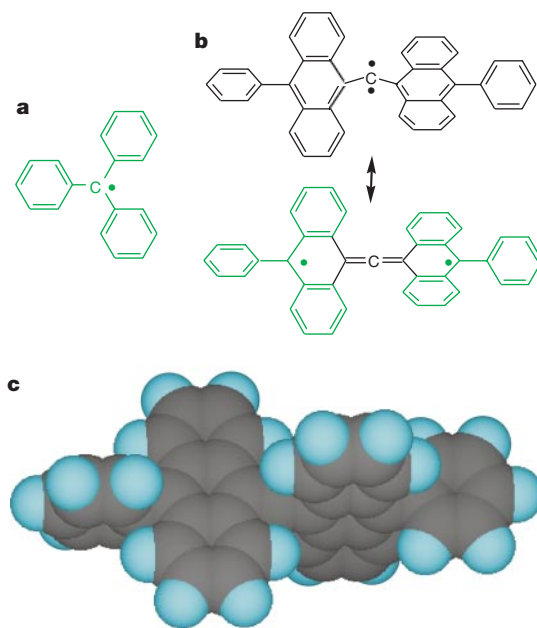


Figure 4 Structure of $^3\mathbf{2b}$ and Gomberg's radical. **a**, Structure of Gomberg's radical. **b**, Resonance structures of $^3\mathbf{2b}$. The lower of these structures can be regarded as two Gomberg radicals connected by an allene bond. **c**, Space-filling model of $^3\mathbf{2b}$, showing how well the carbene centre is shielded by two anthryl groups.

centre (carbene) or localized on two different atoms within the molecule (diradical). One of the best known reactions of triplet carbenes is their interaction with oxygen to form the corresponding ketones, which involves carbonyl oxide formation. In contrast, diradicals react with oxygen to give oxidation products mainly derived from the corresponding peroxides. When **1b** was irradiated in the presence of oxygen, bis [9-(10-phenyl)anthryl]ketone was formed. Laser flash photolysis studies showed the presence of a broad transient absorption at 505 nm ascribable to the carbonyl oxide. Moreover, whereas **3a** aggregated into a trimer, **3b** produced a carbene dimer (**4**) through coupling of two molecules at their carbene centres. These observations, and the fact that we observed the main decay pathway of persistent triplet carbenes in solution to be dimerization, suggest that even though the free electrons in **3b** are extensively delocalized, it is considered more appropriately as a carbene than a diradical.

Although the conceptual ideas underpinning this work were developed more than a century ago, the present findings may still affect modern materials science because triplet carbene units can serve as a useful source of electron spins in high-spin organic molecules that act as models for purely organic ferromagnetics²⁹. To date, the highly transient nature of these species has prevented further development of such systems into usable magnetic materials, but the relatively stable triplet carbene described here may point to new strategies for developing organic ferromagnetic materials. □

Received 5 March; accepted 4 June 2001.

- Gomberg, M. Triphenylmethyl, ein Fall von dreiwertigem Kohlenstoff. *Ber. Deutsch. Chem. Ges.* **33**, 3150–3163 (1900).
- Gomberg, M. An instance of trivalent carbon: Triphenylmethyl. *J. Am. Chem. Soc.* **22**, 757–771 (1900).
- Staudinger, H. & Kupfer, O. Über Reaktionen des Methylens. Diazomethan. *Ber. Deutsch. Chem. Ges.* **45**, 501–509 (1912).
- Hine, J. Carbon dichloride as an intermediate in basic hydrolysis of chloroform. A mechanism for substitution reactions at a saturated carbon atom. *J. Am. Chem. Soc.* **72**, 2438–2445 (1950).
- Doering, W. v. E. & Hoffmann, A. K. Addition of dichlorocarbene to olefins. *J. Am. Chem. Soc.* **76**, 6162–6165 (1954).
- Igaw, A., Grützner, H., Baccareddo, A. & Bertrand, G. Analogous α,α' -bis-carbenoid triply bonded species: Synthesis of a stable λ^3 -phosphinocarbene- λ^3 -phosphaacetylene. *J. Am. Chem. Soc.* **110**, 6463–6466 (1988).
- Arduengo, A. J. III, Harlow, R. L. & Kline, M. A stable crystalline carbene. *J. Am. Chem. Soc.* **113**, 361–363 (1991).
- Bourissou, D., Guerret, O., Gabbai, F. P. & Bertrand, G. Stable carbenes. *Chem. Rev.* **100**, 39–91 (2000).
- Arduengo, A. J. III. Looking for stable carbenes: The difficulty in starting anew. *Acc. Chem. Res.* **32**, 913–921 (1999).
- Regitz, M. Stable carbenes—illusion or reality. *Angew. Chem. Int. Edn Engl.* **30**, 674–676 (1991).
- Dagani, R. Preparation of stable divalent species raises issues of electronic structure. *Chem. Eng. News* **2**, 20–22 (1994).
- Heinemann, C., Müller, T., Apeloig, A. & Schwartz, H. On the question of stability, conjugation, and “aromaticity” in imidazol-2-ylidenes and their silicon analogs. *J. Am. Chem. Soc.* **118**, 2023–2038 (1996).
- Kirmse, W. (ed.) *Carbene Chemistry* 159–504 2nd edn (Academic, New York, 1971).
- Jones, M. & Moss, R. M. (eds) *Carbenes* Vol. II, 159–362 (Wiley, New York, 1975).
- Tomioka, H. Persistent triplet carbenes. *Acc. Chem. Res.* **30**, 315–321 (1997).
- Tomioka, H. Persistent triplet carbenes. *Adv. Carbene Chem.* **2**, 175–214 (1998).
- Hirai, K. & Tomioka, H. A triplet carbene that can almost be bottled. *J. Am. Chem. Soc.* **121**, 10213–10214 (1999).
- Murray, R. W., Trozzolo, A. M., Wasserman, E. & Yager, W. A. E.p.r. of diphenylmethylene, a ground-state triplet. *J. Am. Chem. Soc.* **84**, 3213–3214 (1962).
- Itoh, K. Electron spin resonance of an aromatic hydrocarbon in its quintet ground state. *Chem. Phys. Lett.* **1**, 235–238 (1969).
- Bandon, R. W., Closs, G. & Hutchison, C. A. Jr. Paramagnetic resonance in oriented ground-state triplet molecules. *J. Chem. Phys.* **37**, 1878–1879 (1962).
- Sander, W., Bucher, G. & Wierlacher, S. Carbenes in matrices—spectroscopy, structure, and reactivity. *Chem. Rev.* **93**, 1583–1621 (1993).
- Wasserman, E. *et al.* Electron paramagnetic resonance of 9,9'-dianthrylmethylene. A linear aromatic ground-state triplet methylene. *J. Am. Chem. Soc.* **93**, 6335–6337 (1971).
- Astles, D. S. *et al.* The unusual reactivity of 9,9'-dianthrylcarbene. *J. Org. Chem.* **53**, 6053–6057 (1988).
- Takahashi, Y. *et al.* Triplet di(9-anthryl)carbene undergoes trimerization. *Angew. Chem. Int. Edn Engl.* **39**, 3478–3480 (2000).
- Tomioka, H. Strain and structure of sterically congested triplet carbenes. *Adv. Strained Interesting Organic Mol.* **8**, 83–112 (2000).
- Gibbons, W. A. & Trozzolo, A. M. Spectroscopy and photolysis of a ground-state triplet molecule, diphenylmethylene. *J. Am. Chem. Soc.* **88**, 172–173 (1966).

- Moritani, I., Murahashi, S.-I., Nishino, M., Kimura, K. & Tsubomura, H. Electronic spectra of the products formed by the photolysis of diazo compound at 77 K, possibly identified to carbenes. *Tetrahedr. Lett.* **4**, 373–378 (1966).
- Closs, G., Hutchison, C. A. & Kohler, B. E. Optical absorption spectra of substituted methylenes oriented in single crystals. *J. Chem. Phys.* **44**, 413–414 (1966).
- Iwamura, H. High-spin organic molecules and spin alignment in organic molecular assemblies. *Adv. Phys. Org. Chem.* **26**, 179–253 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank A. Nicolaides for critical reading of the manuscript. This work was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Science, Culture and Sports of Japan, the Nagase Science and Technology Foundation and the Mitsubishi Foundation.

Correspondence and requests for materials should be addressed to H.T. (e-mail: tomioka@chem.mie-u.ac.jp).

Normal faulting in central Tibet since at least 13.5 Myr ago

Peter M. Blisniuk*, Bradley R. Hacker†, Johannes Glodny‡, Lothar Ratschbacher§, Siwen Bi||, Zhenhan Wu||, Michael O. McWilliams¶ & Andy Calvert†

* Institut für Geowissenschaften, Universität Potsdam, D-14415 Potsdam, Germany

† Department of Geological Sciences, University of California, Santa Barbara, California 93106, USA

‡ GeoForschungsZentrum Potsdam, D-14473 Potsdam, Germany

§ Institut für Geologie, Technische Universität Bergakademie Freiberg, D-09596 Freiberg, Germany

|| Institute of Geomechanics, Chinese Academy of Geological Sciences, Beijing 100081, China

¶ Department of Geology and Environmental Sciences, Stanford University, Stanford, California 94305, USA

Tectonic models for the evolution of the Tibetan plateau interpret observed east–west thinning of the upper crust to be the result of either increased potential energy of elevated crust¹ or geodynamic processes that may be unrelated to plateau formation^{2–6}. A key piece of information needed to evaluate these models is the timing of deformation within the plateau. The onset of normal faulting has been estimated to have commenced in southern Tibet between about 14 Myr ago⁷ and about 8 Myr ago⁸ and, in central Tibet, about 4 Myr ago⁹. Here, however, we report a minimum age of approximately 13.5 Myr for the onset of graben formation in central Tibet, based on mineralization ages determined with Rb–Sr and ⁴⁰Ar–³⁹Ar data that post-date a major graben-bounding normal fault. These data, along with evidence for prolonged activity of normal faulting in this and other Tibetan grabens, support models that relate normal faulting to processes occurring beneath the plateau. Thinning of the upper crust is most plausibly the result of potential-energy increases resulting from spatially and temporally heterogeneous changes in thermal structure and density distribution within the crust and upper mantle beneath Tibet. This is supported by recent geophysical and geological data^{10–17}, which indicate that spatial heterogeneity exists in both the Tibetan crust and lithospheric mantle.

The Tibetan plateau consists of several continental fragments (Fig. 1a) that were accreted to the southern margin of Eurasia during the Palaeozoic and Mesozoic eras. Although crustal thickening due to these collisions may have raised portions of the plateau¹⁸, most of the plateau's current elevation is attributed to the India–

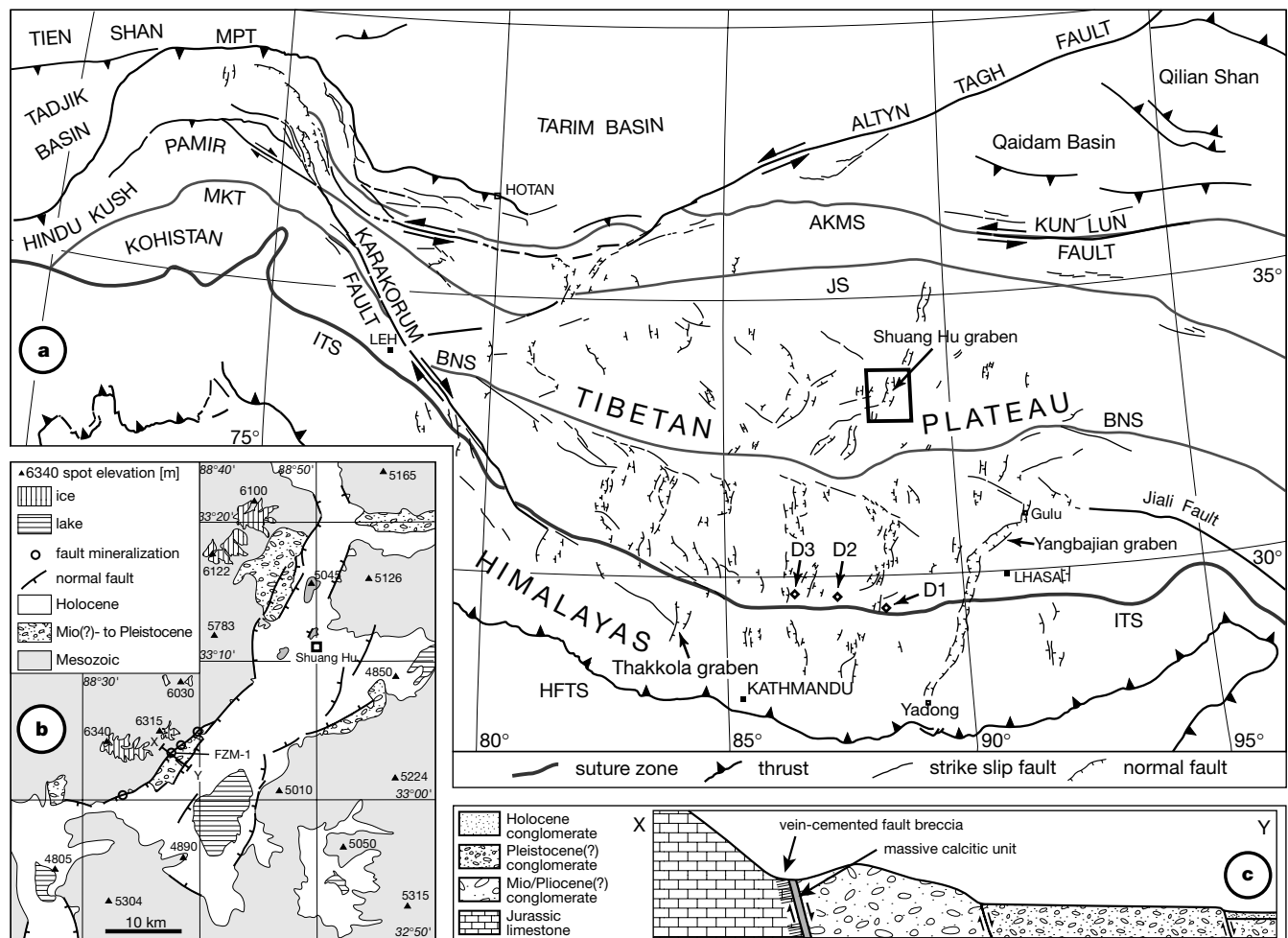


Figure 1 Geological setting of the studied normal fault mineralization. **a**, Generalized tectonic overview map of the Himalaya-Tibet orogen, showing major fault zones with late Cenozoic displacements, and suture zones bordering (and within) the Tibetan plateau. Abbreviations are: AKMS, Ayimaqin-Kunlun-Muttagh suture; BNS, Bangong-Nujiang suture; HFTS, Himalayan frontal thrust system; ITS, Indus-Tsangpo suture; JS, Jinsha suture; MKT, main Karakorum thrust; MPT, main Pamir thrust. D1 indicates location of

dyke dated at ~ 18 Myr ago², D2 and D3 are locations of dykes dated at ~ 18 to ~ 13 Myr ago²². The Lhasa block is located between the ITS and the BNS, the Qiangtang block between the BNS and the JS. **b**, Generalized geological overview map of Shuang Hu graben (for location, see frame in **a**). **c**, Schematic cross-section of normal fault zone at locality FZM-1 along the western margin of Shuang Hu graben (for location, see line X-Y on **b**).

Eurasia collision. Elevation changes cannot be dated directly, but a widely held view is that the onset of normal faulting on the plateau, which extracts potential energy from the crust beneath the plateau, records the time when the maximum elevation of the plateau had been reached¹. A popular hypothesis supporting this view relates uplift of the plateau to convective thinning of its lithospheric root¹⁹. But other models, interpreting normal faulting as a consequence of uplift achieved through other processes, or as unrelated to uplift, are also geodynamically sensible (see refs 2 and 20 for reviews).

Dating the onset of normal faulting on the plateau is a crucial test of the validity of these models, but very few published studies have provided reliable age constraints in this region (Fig. 1a): near the southern margin of the plateau, a minimum age for significant east-west crustal thinning by normal faults of the Thakkola graben is ~ 11 Myr, based on the oldest fault-related sediments observed²¹. Mineralized north-south-trending fractures near the graben dated at ~ 14 Myr indicate at least minor east-west stretching during their formation, and may correlate to normal faulting in the graben⁷. Within the south Tibetan Lhasa block, the onset of significant east-west extension along a low-angle normal fault bounding the Yangbajian graben is dated at 8 ± 1 Myr⁸. However, north-south-trending dykes, about 18–13 Myr old, in the Lhasa block indicate

that at least minor east-west stretching dates back to approximately 18 Myr (refs 2, 22). In the central Tibetan Qiangtang block, the integration of estimates for the Holocene slip rate with the total slip of a major graben-bounding normal fault in the Shuang Hu graben has been used to infer an age of ~ 4 Myr for the initiation of extension⁹.

The data we report here also result from a study of the NNE-trending Shuang Hu graben, ~ 50 km long and 10 km wide, in the central Tibetan Qiangtang block (Fig. 1). The vertical offset along the structurally dominant western graben margin has been estimated at ~ 7 km (ref. 9). The stratigraphy is dominated by Mesozoic strata on the graben shoulders, and Holocene deposits covering the graben floor⁹. The oldest observed deposits interpreted as graben fill are undated coarse conglomerates exposed locally on fault blocks (Fig. 1b, c). Along both graben margins, Quaternary terraces and fans are offset by active normal faults with prominent fault scarps. The geometry of extension in the Shuang Hu graben is constrained by slip directions on minor faults in major graben-bounding normal fault zones striking 020 – 050° and dipping 50 – 80° to the east. Kinematic analysis of these data, to be presented in detail elsewhere, indicates that the orientation of extension is $\sim 120^\circ$. This agrees well with overall structural patterns and fault plane solutions

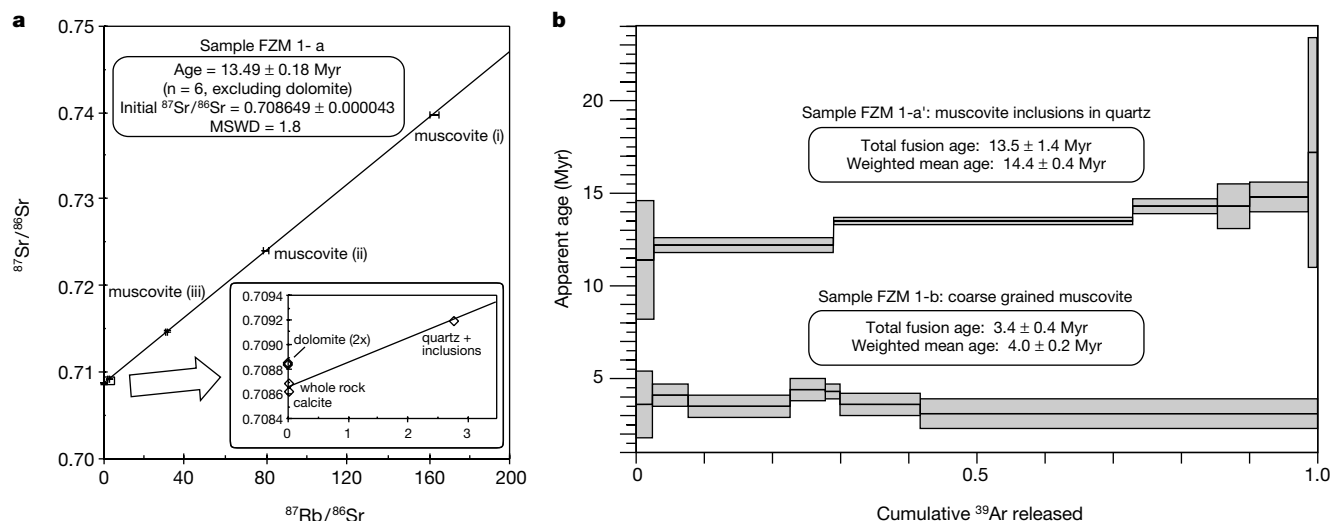


Figure 2 Results of isotope dating. **a**, Rb–Sr data of mineral assemblage from sample FZM 1-a. For meaning of muscovites (i), (ii) and (iii), see Table 1. Dolomite was excluded from the age calculation as it may be secondary. Microtextures suggest that the analysed minerals were not strictly syngenetic, but crystallized in subsequent stages of hydrothermal activity. This is reflected in the isotopic data by a slight non-analytical scatter around the regression line, and by the Sr isotopic disequilibrium between dolomite and calcite. However, integration of the dolomite data in the regression does not change the age value significantly (13.43 ± 0.21 Myr, $n = 8$, mean square of weighted deviates (MSWD) is 24). **b**, Incremental Ar release spectra for muscovite-rich quartz from sample FZM 1-a, and for muscovite from FZM 1-b. Sample FZM 1-a' yielded a sequence of serially increasing step ages, and has a total fusion age of 13.5 ± 1.4 Myr. Three of the

highest temperature steps with the highest K/Ca ratios (26% of the total ^{39}Ar) give a mean age weighted by age uncertainty of 14.4 ± 0.4 Myr, and an inverse isochron age of 13.6 ± 0.9 Myr. The serially increasing step ages may be the result of (1) *in vacuo* inhomogeneous release of ^{40}Ar and ^{39}Ar , in which case the total fusion age of the sample could be considered "best"; (2) two crystallization events, at ~ 14.8 Myr and ~ 12.5 Myr; or (3) a continuum of Ar retention from growth during crystallization at 14.8 to 12.5 Myr. The simplest interpretation, consonant with the Rb–Sr age, is that the total fusion age represents mica crystallization. For muscovite from sample FZM 1-b, the individual laser ages are not internally concordant, low in radiogenic yield, and range from 3.1 to 4.4 Myr. A weighted mean of all seven laser step ages is 4.0 ± 0.2 Myr (MSWD = 2.3). All errors indicated are 2σ .

of earthquakes²³, implying that the Shuang Hu graben is a typical extensional structure of central Tibet.

A minimum age for the onset of important east–west extension in the Shuang Hu graben is recorded by mineralization in the main graben-bounding normal fault zone. At four locations (Fig. 1b) we observed a massive, 2–8-m-thick, calcite-dominated unit along the fault, separating vein-cemented fault breccia of adjacent (probably Jurassic) rocks of the graben shoulder in the west, and (probably Miocene) conglomerates on the fault block to the east (Fig. 1c). At three of the four localities, tufa deposits and mineralized springs occur. At each of the studied sites, the mineralized unit is laterally continuous over tens to hundreds of metres before disappearing under alluvium. Consequently, the mineralized unit appears to be of significant lateral extent, and its emplacement along the fault clearly demonstrates that it post-dates the onset of graben formation.

The fault zone mineralization consists of coarse-grained, non-porous calcite with rare accessory minerals. The investigated rock, from locality FZM-1 (Fig. 1b, c), contains fractures and minor shear planes parallel to the main fault plane, with a typical spacing of ~ 10 cm. Along these, the rock is altered and contains abundant muscovite. Realizing that mineralization may have been dia-

chronous, we separately processed and dated two portions of the sample. One (FZM 1-a) consisted of the unfractured dense portions of the rock. The other (FZM 1-b) contained relatively muscovite-rich bands and rinds associated with the fractures and shear planes.

The Rb–Sr isotope data for muscovite, quartz (with small muscovite, fluid and carbonate inclusions), calcite, and whole rock from sample FZM 1-a yielded an age of 13.5 ± 0.2 Myr (Table 1, Fig. 2a). As an independent test of this result, we performed ^{40}Ar – ^{39}Ar analysis of quartz grains rich in muscovite inclusions from this sample. The apparent age spectrum implies that mineralization occurred at 13.5 ± 1.4 Myr, but a complex crystallization history starting at 14.4 ± 0.4 Myr is possible (Table 2, Fig. 2b). For muscovite from sample FZM 1-b, we obtained a mean ^{40}Ar – ^{39}Ar age of 4.0 ± 0.2 Myr (Table 2, Fig. 2b). We interpret these geochronological data to reflect mineralization at or before 13.5 Myr, and a later stage of muscovite growth or deformation-induced ^{40}Ar loss at ~ 4 Myr.

These results show that normal faulting in the central Tibetan Shuang Hu graben started before ~ 13.5 Myr ago. Furthermore, the 13.5-Myr minimum age of normal faulting, the ~ 4 -Myr age of late-stage mica growth or deformation-induced Ar loss, and significant

Table 1 Rb–Sr data for fault mineralization, Shuang Hu graben

Analysis no.	Material	Rb (p.p.m.)	Sr (p.p.m.)	$^{87}\text{Rb}/^{86}\text{Sr}$	$^{87}\text{Sr}/^{86}\text{Sr}$
PS131	Dolomite (single crystal)	0.35	335.05	0.0030	0.708846 ± 11
PS91	Dolomite (replicate)	0.30	320.41	0.0027	0.708834 ± 13
PS130	Calcite (single crystal)	0.09	795.00	0.0003	0.708618 ± 13
PS54	Muscovite (i)	450.20	8.03	162.6741	0.739691 ± 16
PS57	Muscovite (ii)	443.38	16.12	79.7035	0.724036 ± 12
PS55a	Muscovite (iii)	432.32	39.89	31.3767	0.714629 ± 11
PS53	Quartz + inclusions, leached (iv)	29.63	30.91	2.7733	0.709188 ± 10
PS68	Whole rock	3.74	855.74	0.0126	0.708678 ± 11

These data are from sample FZM 1-a, analysed at GeoForschungsZentrum Potsdam. Muscovite (i), $>100 \mu\text{m}$, treated with acetic acid (25%) then oxalic acid (5%); muscovite (ii), $>100 \mu\text{m}$, treated with acetic acid (25%); muscovite (iii), $>100 \mu\text{m}$, no acid treatment; quartz + inclusions (iv), leached in hot aqua regia (3 h). Errors (2σ) for isochron calculation: $\pm 1.5\%$ for $^{87}\text{Rb}/^{86}\text{Sr}$, $\pm 0.005\%$ for $^{87}\text{Sr}/^{86}\text{Sr}$.

Table 2 ^{40}Ar – ^{39}Ar data for hydrothermal muscovite, Shuang Hu graben

T (°C)	t (min)	^{40}Ar (10^{-15} mol)	$^{40}\text{Ar}/^{39}\text{Ar}$	$^{38}\text{Ar}/^{39}\text{Ar}$	$^{37}\text{Ar}/^{39}\text{Ar}$	$^{36}\text{Ar}/^{39}\text{Ar}$	K/Ca	$\Sigma^{39}\text{Ar}$ (%)	$^{40}\text{Ar}^*$	Age (Myr)
Sample FZM 1-a'										
550	12	8.00	66.1142	6.40×10^{-3}	0.1806	0.2018	2.7	2.58	0.098	11.4 ± 1.6
700	12	18.0	14.4979	0.00	0.4983	0.0255	0.98	28.93	0.481	12.2 ± 0.2
850	12	23.0	11.1971	0.00	0.2978	0.0119	1.6	72.89	0.687	13.5 ± 0.1
950	12	9.10	15.6349	0.00	0.3414	0.0252	1.4	85.27	0.524	14.3 ± 0.2
1,000	12	3.20	14.1738	0.00	0.2807	0.0204	1.7	90.00	0.574	14.3 ± 0.6
1,075	12	8.20	19.8975	0.00	0.2336	0.0388	2.1	98.71	0.424	14.8 ± 0.4
1,150	12	5.40	96.4782	0.00	1.7430	0.2931	0.28	99.90	0.102	17.2 ± 3.1
1,225	12	7.30	2130.2571	7.50×10^{-2}	23.6827	7.0862	0.021	99.97	0.017	62.7 ± 369.8
1,300	12	6.70	4751.1260	0.00	19.5559	16.0119	0.025	100.00	0.004	34.2 ± 1959.9
Sample FZM 1-b										
	9.7	54.9376	ND	0.2165	0.1297	2.3	2.39	0.302	3.6 $\pm$ 0.9	
	21.0	52.6670	ND	0.1414	0.1149	3.5	7.64	0.355	4.1 $\pm$ 0.3	
	58.0	52.2451	ND	0.0801	0.1231	6.1	22.61	0.304	3.5 $\pm$ 0.3	
	19.0	50.5137	ND	0.1441	0.1038	3.4	27.75	0.392	4.4 $\pm$ 0.3	
	8.4	52.4667	ND	0.2613	0.1116	1.9	29.92	0.371	4.3 $\pm$ 0.2	
	46.0	52.0011	ND	0.0921	0.1210	5.3	41.73	0.312	3.6 $\pm$ 0.3	
	210.0	48.0376	ND	0.1395	0.1142	3.5	100.00	0.298	3.1 $\pm$ 0.4	

Sample FZM 1-a': quartz grains rich in muscovite inclusions, hand-picked from the mineral separate used for Rb–Sr analysis, and analysed by step heating in a resistance furnace (UCSB). $J = 0.0009744$. Sample FZM 1-b: muscovite hand-picked from crushed rock and analysed by laser step-heating of groups of grains (Stanford University). $J = 0.0001216$ and $0.001218 \pm 0.5\%$. J , irradiation flux parameter; T , temperature of furnace; t , time in furnace at temperature T ; ^{40}Ar , moles of ^{40}Ar corrected for blank and reactor-produced ^{40}Ar ; $\Sigma^{39}\text{Ar}$, cumulative ^{39}Ar released; $^{40}\text{Ar}^*$, radiogenic fraction; ND, not determined. Ratios are corrected for blanks, decay, and interference; errors on age are $\pm 1\sigma$.

Quaternary normal faulting demonstrate prolonged upper-crustal thinning in that region. This is similar to the situation farther south on the Tibetan plateau. In the south Tibetan Lhasa block, north–south-trending dyke swarms indicate east–west stretching since ~ 18 Myr ago²², and studies in the Yangbajian graben, formed ~ 8 Myr ago⁸, have documented Quaternary and recent normal faulting³. In the Thakkola graben, normal faulting occurred in multiple phases from ~ 11 Myr into the Quaternary^{24,25}, and may have started as early as ~ 14 Myr ago⁷.

The available age data, although not sufficient to resolve the chronology of Tibet's upper-crustal extension in detail, call into question the relevance of models relating Tibetan normal faulting to processes occurring near the margins of (or outside) the plateau. For example, mechanisms attributing normal faulting in southern Tibet to motion along the Karakorum–Jiali fault zone³, to radial convergence⁴, or to oroclinal bending^{5,6} are not likely to have triggered significant extension in central Tibet, although they may have influenced its distribution in southern Tibet. A link between crustal thinning in Tibet and important regional normal faulting in north-central Asia (for example, Shanxi, Baikal)²—poorly dated, but thought to have initiated during latest Miocene to Early Pliocene times^{26,27}—is also not supported by evidence for normal faulting in central and southern Tibet at or before 13.5 Myr ago. In contrast, the view that the onset of crustal thinning is related to increased potential energy of elevated crust is consistent with this evidence, and supported by palaeoelevation estimates for ~ 11 -Myr-old Thakkola graben sediments that are similar to modern elevations on the plateau²¹.

The most plausible causes for important potential-energy increases are changes in the thermal structure and density distribution within the crust and upper mantle beneath Tibet. Geophysical and geological evidence^{10–17} indicates a heterogeneous crust and lithospheric mantle structure beneath Tibet—characterized, for example, by decreasing crustal thickness from south to north, hot upper crust but cold lower crust and lithospheric mantle in the south, and hot lower crust and lithospheric mantle in the north. This makes it difficult to interpret increased potential energy as being caused by the same process occurring simultaneously beneath the entire plateau, as has been suggested for convective removal of lower lithosphere¹⁹. It is more likely that the processes controlling potential-energy changes within the crust and upper mantle beneath Tibet have been heterogeneous spatially and temporally. In the interpretation that crustal thinning in Tibet is the result of increased potential energy of elevated crust, the few available age constraints on normal faulting imply high surface elevation in parts

of southern and central Tibet, but not necessarily throughout the entire present-day plateau, by ~ 14 Myr ago. If important regional climate changes at ~ 8 Myr ago^{28–30} were related to the evolution of the Tibetan plateau, they probably occurred after the plateau had attained a critical minimum elevation and/or size. It seems possible that the value of either one or both of these parameters increased between ~ 14 Myr and ~ 8 Myr ago. □

Received 30 January; accepted 22 June 2001.

- Molnar, P. & Tapponnier, P. Active tectonics of Tibet. *J. Geophys. Res.* **83**, 5361–5375 (1978).
- Yin, A. & Harrison, T. M. Geologic evolution of the Himalayan–Tibetan orogen. *Annu. Rev. Earth Planet. Sci. Lett.* **28**, 211–280 (2000).
- Armijo, R., Tapponnier, P. & Han, T. Late Cenozoic strike-slip faulting in southern Tibet. *J. Geophys. Res.* **94**, 2787–2838 (1989).
- Klootwijk, C. T., Conaghan, P. J. & Powell, C. M. The Himalayan arc: large-scale continental subduction, oroclinal bending and back-arc spreading. *Earth Planet. Sci. Lett.* **75**, 167–183 (1985).
- McCaffrey, R. & Nabelek, J. Role of oblique convergence in the active deformation of the Himalayas and southern Tibet plateau. *Geology* **26**, 691–694 (1998).
- Ratschbacher, L., Frisch, W., Lui, G. & Chen, C. Distributed deformation in southern and western Tibet during and after the India–Asia collision. *J. Geophys. Res.* **99**, 19917–19945 (1994).
- Coleman, M. & Hodges, K. Evidence for Tibetan plateau uplift before 14 Myr ago from a new minimum age for east–west extension. *Nature* **374**, 49–52 (1995).
- Harrison, T. M., Copeland, P., Kidd, W. S. F. & Lovera, O. M. Activation of the Nyaingentanghla Shear Zone: Implications for uplift of the southern Tibetan Plateau. *Tectonics* **14**, 658–676 (1995).
- Yin, A. *et al.* Evidence for significant Late Cenozoic E–W extension in North Tibet. *Geology* **27**, 787–790 (1999).
- Nelson, K. D. *et al.* An INDEPTH view of the structure of the lithosphere beneath southern Tibet. *Science* **274**, 1684–1688 (1996).
- Owens, T. J. & Zandt, G. Implications of crustal property variations for models of Tibetan plateau evolution. *Nature* **387**, 37–43 (1997).
- Kosarev, G. *et al.* Seismic evidence for a detached Indian lithospheric mantle beneath Tibet. *Science* **283**, 1306–1309 (1999).
- Hacker, B. R. *et al.* Hot and dry deep crustal xenoliths from Tibet. *Science* **287**, 2463–2466 (2000).
- Huang, W.-C. *et al.* Seismic polarization anisotropy beneath the central Tibetan Plateau. *J. Geophys. Res.* **105**, 27979–27989 (2000).
- Wei, W. *et al.* Detection of widespread fluids in the Tibetan crust by magnetotelluric studies. *Science* **292**, 716–718 (2001).
- Zhao, W. *et al.* Crustal structure of central Tibet as derived from project INDEPTH wide-angle seismic data. *Geophys. J. Int.* **145**, 486–498 (2001).
- Rodgers, A. J. & Schwartz, S. Y. Lithospheric structure of the Qiangtang Terrane, northern Tibetan Plateau, from complete regional waveform modeling: Evidence for partial melt. *J. Geophys. Res.* **103**, 7137–7152 (1998).
- England, P. & Searle, M. The Cretaceous–Tertiary deformation of the Lhasa Block and its implications for crustal thickening in Tibet. *Tectonics* **5**, 1–14 (1986).
- England, P. C. & Houseman, G. A. Extension during continental convergence, with application to the Tibetan plateau. *J. Geophys. Res.* **94**, 17561–17579 (1989).
- Harrison, T. M., Copeland, P., Kidd, W. S. F. & Yin, A. Raising Tibet. *Science* **255**, 1663–1670 (1992).
- Garzione, C. N., Dettman, D. L., Quade, J., DeCelles, P. G. & Butler, R. F. High times on the Tibetan Plateau: Paleoelevation of the Thakkola graben, Nepal. *Geology* **28**, 339–342 (2000).
- Williams, H., Turner, S., Kelley, S. & Harris, N. Age and composition of dikes in Southern Tibet: New constraints on the timing of east–west extension and its relationship to postcollisional volcanism. *Geology* **29**, 339–342 (2001).

23. Molnar, P. & Lyon-Caen, H. Fault plane solutions of earthquakes and active tectonics of the northern and eastern parts of the Tibetan Plateau. *Geophys. J. Int.* **99**, 123–153 (1989).
24. Hodges, K. V. *et al.* Simultaneous Miocene extension and shortening in the Himalayan orogen. *Science* **258**, 1466–1470 (1992).
25. Hurtado, J. M., Hodges, K. V. & Whipple, K. X. Neotectonics of the Thakkola graben and implications for recent activity on the South Tibetan fault system in the central Nepal Himalaya. *Bull. Geol. Soc. Am.* **113**, 222–240 (2001).
26. Zhang, Y., Vergely, P. & Mercier, J. L. Pliocene-Quaternary faulting pattern and left-slip propagation tectonics in North China. *Episodes* **22**, 84–88 (1999).
27. Logatchev, N. A. & Zorin, Y. A. Evidence and causes of the two-stage development of the Baikal rift. *Tectonophysics* **143**, 225–234 (1987).
28. Kroon, D., Steens, T. & Troelstra, S. R. Onset of monsoonal related upwelling in the western Arabian Sea as revealed by planktonic foraminifers. *Proc. ODP Sci. Res.* **116**, 257–263 (1991).
29. Quade, J., Cerling, T. E. & Bowman, J. R. Development of Asian monsoon revealed by marked ecological shift during the latest Miocene in northern Pakistan. *Nature* **342**, 163–166 (1989).
30. An, Z., Kutzbach, J. E., Prell, W. L. & Porter, S. C. Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since Late Miocene times. *Nature* **411**, 62–66 (2001).

Acknowledgements

We thank E. Gnos, M. Strecker, W. Kidd, A. Yin and M. Edwards for discussions, and P. Molnar for comments and suggestions. This work was supported by Deutsche Forschungsgemeinschaft, the US NSF, and the Chinese Academy of Geological Sciences.

Correspondence and requests for materials should be addressed to P.M.B. (e-mail: blisniuk@rz.uni-potsdam.de).

Resistance to mantle flow inferred from the electromagnetic strike of the Australian upper mantle

Fiona Simpson

Geophysics Institute, University of Göttingen, Germany, and Institute of Geological & Nuclear Sciences, 69 Gracefield Road, PO Box 30-368, Lower Hutt, New Zealand

Seismic anisotropy is thought to result from the strain-induced lattice-preferred orientation of mantle minerals, especially olivine^{1,2}, owing to shear waves propagating faster along the *a*-axis of olivine crystals than along the other axes. This anisotropy results in birefringence, or ‘shear-wave splitting’³, which has been investigated in numerous studies^{1,4}. Although olivine is also anisotropic with respect to electrical conductivity⁵ (with the *a*-axis being most conductive), few studies of the electrical anisotropy of the upper mantle have been undertaken, and these have been limited to relatively shallow depths in the lithospheric upper mantle^{6,7}. Theoretical models of mantle flow have been used to infer that, for progressive simple shear imparted by the motion of an overriding tectonic plate, the *a*-axes of olivine crystals should align themselves parallel to the direction of plate motion^{8,9}. Here, however, we show that a significant discrepancy exists between the electromagnetic strike of the mantle below Australia and the direction of present-day absolute plate motion¹⁰. We infer from this discrepancy that the *a*-axes of olivine crystals are not aligned with the direction of the present-day plate motion of Australia, indicating resistance to deformation of the mantle by plate motion.

Most studies of continental seismic anisotropy have been founded on analyses of the splitting of SKS shear waves. A disadvantage of this is the ambiguity associated with the depth at which anisotropy occurs. This has created a controversy, with some authors attributing seismic anisotropy to ‘frozen-in’ lithospheric anisotropy associated with palaeo-orogenic deformations^{4,11}, and others attributing it to sublithospheric anisotropy owing to strain-induced lattice-preferred orientation of olivine¹. This ambiguity can be resolved to

some extent using magnetotelluric sounding, as the depth to an electrically conducting, anisotropic layer can be well constrained, whereas the thickness of the layer cannot. The magnetotelluric technique is a passive technique that involves measuring fluctuations in the natural electric (*E*) and magnetic (*H*) fields in orthogonal directions at the surface of the Earth. The orthogonal components of the horizontal *E* and *H* fields are related via a complex impedance tensor, *Z*:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix} \quad (1)$$

For a one-dimensional or two-dimensional Earth with *x* or *y* aligned along strike (direction of maximum conductance), the diagonal elements *Z*_{xx} and *Z*_{yy} are zero. Mathematically, a one-dimensional anisotropic Earth is equivalent to a two-dimensional Earth. Measured data rarely have zero diagonal impedance elements in any coordinate system, but can often be described by a ‘superimposition (decomposition) model’ in which the data are decomposed into a local, non-inductive response (commonly termed ‘galvanic’) owing to multi-dimensional heterogeneities with dimensions significantly less than the inductive scale length of the data and a regional inductive response to an underlying two-dimensional structure^{12,13}. For data aligned in the (*x*′, *y*′) coordinate system of the regional

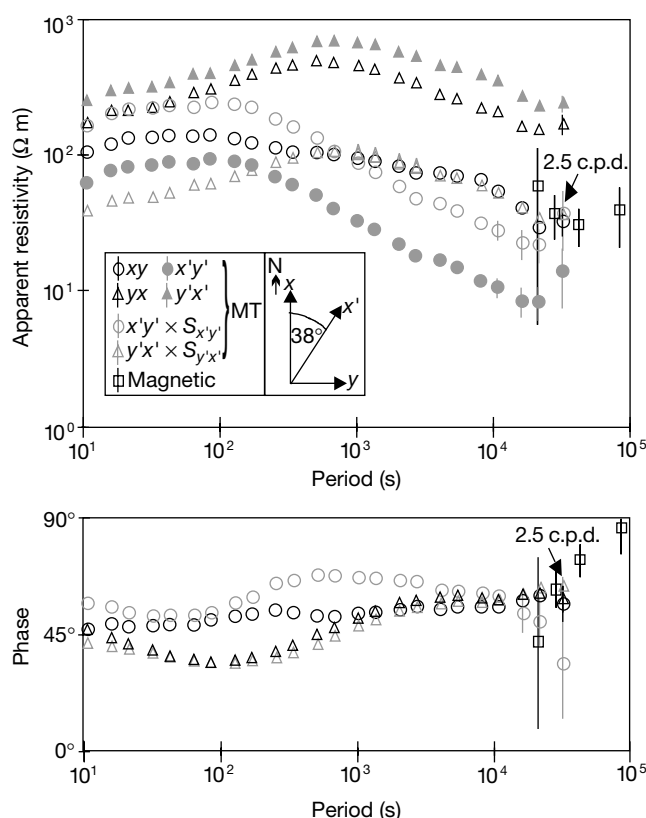


Figure 1 Apparent resistivity and impedance phases at station ASP before (*x*, *y*) and after rotation (*x*′, *y*′) and correction for static shifts. The magnetotelluric apparent resistivities are corrected for static shift by applying scaling factors (*S*_{*x*′*y*′}, *S*_{*y*′*x*′}) such that the magnetotelluric data at 2.5 c.p.d. correspond to the level of the magnetic transfer functions. Because the horizontal magnetic field induced by solar quiet (Sq) is strongly polarized in the east–west (*y*) direction, magnetic transfer functions can only be derived for the polarization corresponding to the *xy* polarization of the magnetotelluric data. There is, however, good correspondence between the *yx* magnetotelluric phases and the Sq phases. The phases are not affected by static shift, but are changed by rotation.

two-dimensional structure, the impedance tensor can then be expanded as:

$$Z = \begin{vmatrix} a_{11} & a_{12} & 0 & Z_{x'y'} \\ a_{21} & a_{22} & -Z_{y'x'} & 0 \end{vmatrix} = \begin{vmatrix} -a_{12}Z_{y'x'} & a_{11}Z_{x'y'} \\ -a_{22}Z_{y'x'} & a_{21}Z_{x'y'} \end{vmatrix} \quad (2)$$

$$\text{or } Z = AZ_{2D}$$

Within each column only one phase occurs because the assumption of galvanic distortion requires that the elements of the distortion matrix, A , must be real and frequency-independent. The condition that the phases of the two elements of the impedance tensor pertaining to the same electric field vector should be equal is used to find the regional strike angle (α) from

$$Z = R_\alpha AZ_{2D}R_\alpha^T \quad (3)$$

where R is the rotation matrix and superscript T denotes transpose. The regional strike is determined by rotating the measured impedance to find the best fit to the model represented by equation (2). Errors in the electromagnetic strike are computed as described in the Methods.

Because electromagnetic waves propagate diffusively, short-period waves are attenuated by conductive structures in the crust. Therefore, a pre-requisite for studying electrical anisotropy of the upper mantle is the availability of high-quality, long-period magnetotelluric transfer functions. Impedance tensors have been obtained for a period range spanning $10\text{--}10^5$ s (Fig. 1) from four stations lying $150\text{--}300$ km apart on the North Central craton of Australia (Fig. 2), far away from known tectonic boundaries and coastlines. Consideration of the distortion parameters indicates that the data can be described by Class 5 (a regional two-dimensional anomaly with strong local distortion) of Bahr's classification of distortion types¹⁴. Therefore, the decomposition hypothesis is an appropriate model for determining the electromagnetic strike of these data.

Three-dimensional conductivity anomalies significantly shallower than the target depth also cause local distortion (static shift) of the amplitudes of the electric fields. Static shift causes the impedance magnitudes to be shifted by real scaling factors, and, if

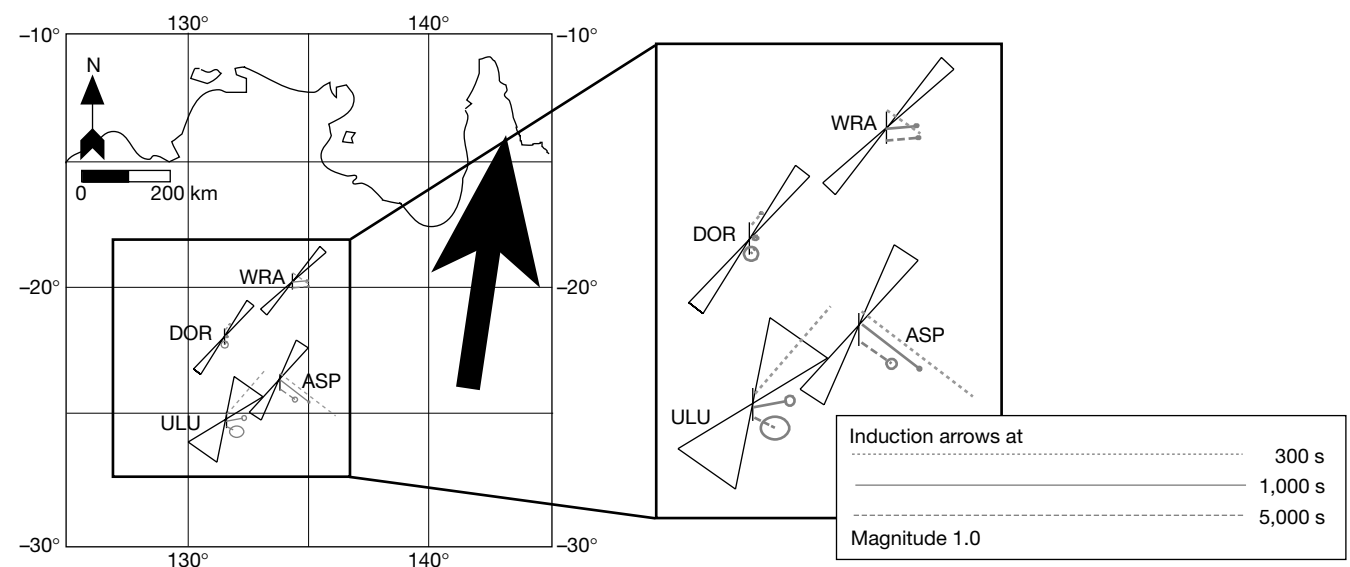


Figure 2 Map showing locations of magnetotelluric stations. Fans at station locations indicate electromagnetic strike directions for the period range $10^3\text{--}10^4$ s, corresponding to depths deeper than 150 km. Induction arrows and their error ellipses are marked in grey for periods of 300 s, $1,000$ s and $5,000$ s. Bold arrow indicates the direction of present-day plate motion. The electromagnetic strike deviates significantly from the

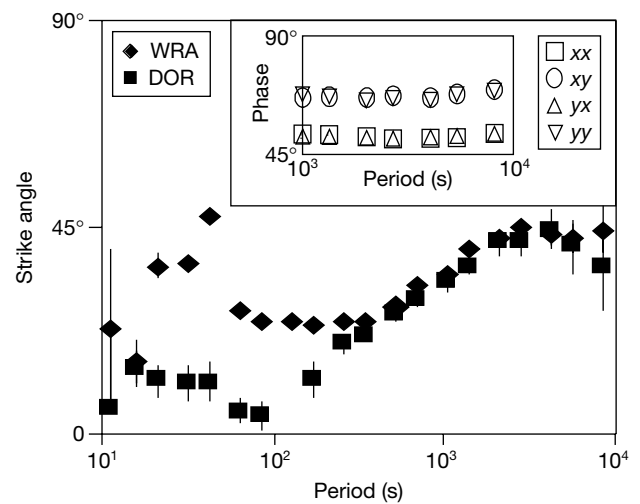


Figure 3 Phase-sensitive electromagnetic strike for two stations, plotted as a function of period. Errors are derived from equation (5). Inset shows the impedance phases for all four elements of the impedance tensor in the period range $10^3\text{--}10^4$ s at DOR following rotation to the phase-sensitive strike. This demonstrates that the condition expressed in equation (2) is fulfilled: the phases belonging to a single column of the impedance tensor are equal.

not removed, results in incorrect estimation of the depth and conductance of deeper conductivity anomalies. The scaling factor cannot be determined uniquely from equation (2). This non-uniqueness problem can be overcome by using an estimate of the impedance that is determined from the magnetic fields alone, and is, therefore, independent of the distorted electric field. Static shifts have been corrected using magnetic transfer functions derived from solar quiet (Sq) variations^{15,16} (daily harmonic fluctuations of the Earth's magnetic field generated by the solar quiet current vortex in the ionosphere). The impedances at 2.5 cycles per day (c.p.d.) have been scaled to give the inductive scale lengths determined by interpolation of the ratios of vertical to horizontal magnetic fields

direction of plate motion. The small induction arrows support the conclusion that the electromagnetic strike arises through anisotropy at upper mantle depths by demonstrating the absence of significant vertical magnetic fields, which would be generated if lateral conductivity gradients were present.

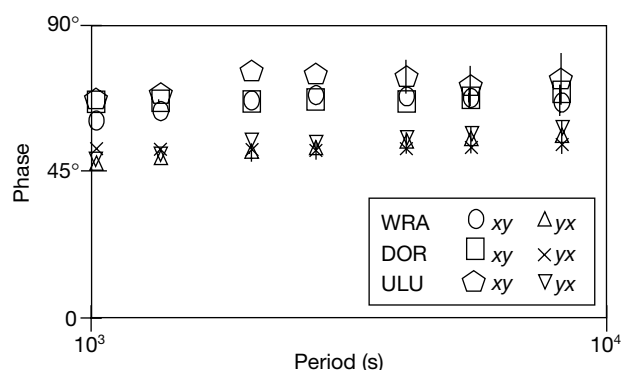


Figure 4 Impedance phases in the period range 10^3 – 10^4 s for three sites after rotation to the phase-sensitive strike. The phase splitting is roughly the same at all sites. This identical phase splitting without accompanying vertical magnetic fields (Fig. 2) fortifies the conclusion that the strike of roughly 38° is generated by deep anisotropy rather than by lateral structure. Vertical lines are error bars. At the longest period there is one error bar belonging to a hexagon and one belonging to a circle. At other periods, error bars belong to hexagons.

at periods of 2–3 c.p.d. (Fig. 1). Interpolation is necessary, because the magnetic transfer functions are computed at the frequencies of the Sq harmonics, whereas the magnetotelluric transfer functions must be computed at frequencies in the continuum between the Sq spectral lines to avoid contamination from the non-uniform source field. Because the Sq current vortex flows roughly north–south at the latitude of the survey, the induced magnetic field is strongly polarized in the east–west direction. Therefore independent determinations of the static shift for the two polarizations of the impedance tensor cannot be obtained. Convergence of the impedance phases at long periods indicates, however, that a one-dimensional static shift correction relying on B_z/B_y is adequate.

The errors in the impedance phases are less than 1.5° throughout the 10 – 10^4 s period range, which corresponds to penetration depths of about 10–300 km, and significant phase splitting occurs. This combination of large impedance phase splits and low measurement errors enables a well-constrained phase-sensitive (electromagnetic) strike to be calculated. The electromagnetic strike is shown as a function of period for two sites in Fig. 3. Equivalence is achieved between the phases of Z_{xx} and Z_{yy} , and Z_{yx} and Z_{xy} , respectively (in other words, the condition that the elements of $\mathbf{A}$ are real is met) following rotation to the phase-sensitive strike for the period range 10^3 – 10^4 s (Fig. 3).

Between 10^3 and 10^4 s, all sites indicate a consistent electromagnetic strike of $38 \pm 13^\circ$ ($38 \pm 7^\circ$ omitting the ULU site, where the strike is less well constrained; Fig. 2). Figure 2 also shows induction arrows at 300 s, 1,000 s and 5,000 s. Induction arrows are graphical representations of the ratio of vertical to horizontal magnetic field components depicted as vectors that point towards anomalous internal concentrations of current¹⁷. Vertical magnetic fields are generated by lateral conductivity gradients. Therefore, induction arrows indicate lateral variations in conductivity. At periods longer than 1,000 s, the induction arrow magnitudes are roughly zero at all sites. Induction arrow magnitudes are negligible at all periods at DOR and WRA. Lateral structure is very unlikely to generate a large split in the impedance phases at a single site with no associated vertical magnetic field. Also, if the splitting observed between the two polarizations of the impedance phases were caused by lateral gradients, then the amount of phase splitting should change with distance from the lateral structure responsible. The impedance phases for the three sites shown in Fig. 4 are remarkably similar in the period range 10^3 – 10^4 s. The absence of a vertical magnetic field and consistent phase splitting over geographical distances greater than 300 km strongly support the conclusion that the 38°

strike is generated by deep anisotropy rather than by lateral structure¹⁸.

One-dimensional modelling using the Montecarlo technique was done to investigate the range of anisotropic conductance models that can fit the data. An anisotropic layer, corresponding to the period range 10^3 – 10^4 s, is required deeper than 150 km, with ratios of strike to cross-strike resistivities of about 2–3. These conclusions are robust despite uncertainties in static shifts (Fig. 1).

The electromagnetic strike can be influenced by macroscopic structures¹⁹ (such as shear zones) preserved from palaeo-orogenic deformation. But grain-scale anisotropy is the more likely cause of the direction-dependent enhancement of conductivity at depths exceeding 150 km, because at the temperatures expected deeper than 150 km, macroscopic structural lineaments are unlikely to be preserved.

Deeper than 150 km, the electromagnetic strike does not match the present-day direction of absolute plate motion, which is $N9.6^\circ E$, but may instead relate to a past trajectory of plate motion. In the region studied, SV-wave azimuthal anisotropy at a depth of 150 km (constrained by dispersion of Rayleigh waves) is also not aligned with the present-day plate motion direction²⁰. The electromagnetic strike direction corresponds well with the fast direction of S-wave splitting at 200 km (ref. 21). The directions of seismic and electrical anisotropy indicate that below the North Central craton of Australia, strains associated with present-day plate motion have not dominated mantle deformation sufficiently to induce alignment of olivine in the direction of plate motion. This raises an intriguing paradox given that, moving with a speed of $8.44 \text{ cm year}^{-1}$, the Australian plate is one of the fastest-moving plates on Earth¹⁰, and should impart a high strain on the mantle. This indicates recent resistance to deformation of the mantle by plate motion, possibly owing to a cold, high-viscosity layer in the base of the lithospheric mantle or sublithospheric mantle below central Australia. $\square$

Methods

Here, I extend Bahr's original formulation¹² to include error propagation from the measured impedance tensor to the phase-sensitive regional strike (α):

$$\alpha = \frac{1}{2} \arctan \left(\frac{[S_1, S_2] - [D_1, D_2]}{[S_1, D_1] + [S_2, D_2]} \right) \quad (4)$$

where $S_1 = Z_{xx} + Z_{yy}$, $S_2 = Z_{yy} + Z_{xx}$, $D_1 = Z_{xx} - Z_{yy}$, $D_2 = Z_{yy} - Z_{xx}$ and $[Z_1, Z_2] = \text{Re}Z_1 \text{Im}Z_2 - \text{Re}Z_2 \text{Im}Z_1$. The following error ($d\alpha$) incorporates both the errors in the measured data and the degree of phase splitting:

$$d\alpha = \frac{1}{2} \frac{1}{1 + (\tan 2\alpha)^2} d(\tan 2\alpha) \quad (5)$$

where

$$d(\tan 2\alpha) = \sqrt{\left(\frac{d([S_1, S_2] - [D_1, D_2])}{[S_1, D_1] + [S_2, D_2]} \right)^2 + \left(\frac{d([S_1, D_1] + [S_2, D_2])([S_1, S_2] - [D_1, D_2])}{([S_1, D_1] + [S_2, D_2])^2} \right)^2}$$

and $d([S_1, S_2] - [D_1, D_2])^2 = \Delta S_1^2 S_2^2 + S_1^2 \Delta S_2^2 + \Delta D_1^2 D_2^2 + D_1^2 \Delta D_2^2$ and $d([S_1, D_1] + [S_2, D_2])^2 = \Delta S_1^2 D_1^2 + S_1^2 \Delta D_1^2 + \Delta S_2^2 D_2^2 + S_2^2 \Delta D_2^2$.

Received 24 April; accepted 22 June 2001.

- Vinnik, L. P., Makeyeva, L. I., Milev, A., Usenko, A. Y. Global patterns of azimuthal anisotropy and deformations in the continental mantle. *Geophys. J. Int.* **111**, 433–447 (1992).
- Vinnik, L. P., Green, R. W. E. & Nicolaysen, L. O. Recent deformations of the deep continental root beneath southern Africa. *Nature* **375**, 50–52 (1995).
- Christensen, N. I. The magnitude, symmetry, and origin of upper mantle anisotropy based on fabric analyses of ultramafic tectonites. *Geophys. J. R. Astron. Soc.* **76**, 89–112 (1984).
- Silver, P. G. & Chan, W. W. Shear wave splitting and subcontinental mantle deformation. *J. Geophys. Res.* **96**, 16429–16454 (1991).
- Mackwell, S. J. & Kohlstedt, D. L. Diffusion of hydrogen in olivine: Implications for water in the mantle. *J. Geophys. Res.* **95**, 5079–5088 (1990).
- Kurtz, R. D., Craven, J. A., Niblett, E. R. & Stevens, R. The conductivity of the crust and mantle beneath the Kapuskasing Uplift: Electrical anisotropy in the upper mantle. *Geophys. J. Int.* **113**, 483–498 (1993).
- Mareschal, M. et al. Archaean cratonic roots, mantle shear zones and deep electrical anisotropy. *Nature* **375**, 134–137 (1995).

8. McKenzie, D. Finite deformation during fluid flow. *Geophys. J. R. Astron. Soc.* **58**, 689–715 (1979).
9. Ribe, N. M. Seismic anisotropy and mantle flow. *J. Geophys. Res.* **94**, 4213–4223 (1989).
10. Gripp, A. E. & Gordon, R. G. Current plate velocities relative to the hotspots incorporating the NUVEL-1 global plate motion model. *Geophys. Res. Lett.* **17**, 1109–1112 (1990).
11. Barruol, G., Silver, P. G. & Vauchez, A. Seismic anisotropy in the eastern United States: Deep structure of a complex continental plate. *J. Geophys. Res.* **102**, 8329–8348 (1997).
12. Bahr, K. Interpretation of the magnetotelluric impedance tensor: regional induction and local telluric distortion. *J. Geophys.* **62**, 119–127 (1988).
13. Groom, R. W. & Bailey, R. C. Decomposition of the magnetotelluric impedance tensor in the presence of local three-dimensional galvanic distortion. *J. Geophys. Res.* **94**, 1913–1925 (1989).
14. Bahr, K. Geological noise in magnetotelluric data: a classification of distortion types. *Phys. Earth Planet. Inter.* **66**, 24–38 (1991).
15. Schmucker, U. J. An introduction to induction anomalies. *Geomagn. Geoelectr.* **22**, 9–33 (1970).
16. Bahr, K. & Filloux, J. H. Local Sq response functions from EMTSLAB data. *J. Geophys. Res.* **94**, 14195–14200 (1989).
17. Parkinson, W. D. Directions of rapid geomagnetic fluctuations. *Geophys. J. R. Astron. Soc.* **2**, 1–14 (1959).
18. Bahr, K. & Duba, A. Is the asthenosphere electrically anisotropic? *Earth Planet. Sci. Lett.* **178**, 87–95 (2000).
19. Simpson, F. A three-dimensional electromagnetic model of the southern Kenya Rift: departure from two-dimensionality as a consequence of a rotating stress field. *J. Geophys. Res.* **105**, 19321–19334 (2000).
20. Debayle, E. SV-wave azimuthal anisotropy in the Australian upper mantle: preliminary results from automated Rayleigh waveform inversion. *Geophys. J. Int.* **137**, 747–754 (1999).
21. Debayle, E. & Kennett, B. L. N. The Australian continental upper mantle: structure and deformation inferred from surface waves. *J. Geophys. Res.* **105**, 25423–25450 (2000).

Acknowledgements

This research was supported by the Indian Institute of Geomagnetism, Bombay and the Geophysics Institute, University of Göttingen, via an initiative of Baldev Arora. I thank K. Bahr, A. Gatzemeier and P. Haehy for help with fieldwork, G. Heinson for hospitality during my sojourn at the University of Adelaide and J. Booker for comments on an earlier version of this paper.

Correspondence and requests for materials should be addressed to F.S. (e-mail: f.simpson@gns.cri.nz).

Unicellular cyanobacteria fix N₂ in the subtropical North Pacific Ocean

Jonathan P. Zehr*, John B. Waterbury†, Patricia J. Turner*, Joseph P. Montoya‡, Enoma Omoregie*, Grieg F. Steward*, Andrew Hansen§ & David M. Karl§

*Department of Ocean Sciences and Institute of Marine Sciences, University of California, Santa Cruz, California 95064, USA

†Biology Department, Woods Hole Oceanographic Institute, Woods Hole, Massachusetts 02543, USA

‡School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

§School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA

Fixed nitrogen (N) often limits the growth of organisms in terrestrial and aquatic biomes^{1,2}, and N availability has been important in controlling the CO₂ balance of modern and ancient oceans^{3,4}. The fixation of atmospheric dinitrogen gas (N₂) to ammonia is catalysed by nitrogenase and provides a fixed N for N-limited environments^{2,5}. The filamentous cyanobacterium *Trichodesmium* has been assumed to be the predominant oceanic N₂-fixing microorganism since the discovery of N₂ fixation in *Trichodesmium* in 1961 (ref. 6). Attention has recently focused on oceanic N₂ fixation because nitrogen availability is generally limiting in many oceans, and attempts to constrain the global atmosphere–ocean fluxes of CO₂ are based on basin-scale N balances^{7–9}. Biogeochemical studies and models have suggested that total N₂-fixation rates may be substantially greater than previously believed^{7,8} but cannot be reconciled with observed *Trichodesmium* abundances^{8,9}. It is curious that there are so few

known N₂-fixing microorganisms in oligotrophic oceans when it is clearly ecologically advantageous. Here we show that there are unicellular cyanobacteria in the open ocean that are expressing nitrogenase, and are abundant enough to potentially have a significant role in N dynamics.

Water samples were collected during several cruises to the Hawaii Ocean Time-series (HOT) station ALOHA to determine whether N₂-fixing microorganisms were present and, if so, expressing nitrogenase. Previous results indicated that diverse diazotrophs, potentially including unicellular cyanobacteria, were present at station ALOHA based on the analysis of nitrogenase genes (specifically *nifH*, which encodes the Fe protein component)^{10,11}. In our study, RNA was extracted to determine if these microorganisms were expressing *nifH* under *in situ* conditions, which would indicate that they are responding to conditions of limited N and synthesizing the nitrogenase protein. Also, we looked for the putative unicellular cyanobacteria by microscopy, cultivated isolates and performed experiments to see if the microorganisms could fix N₂ both *in situ* and in culture.

Nitrogenase gene transcripts (messenger RNA) attributable to organisms other than *Trichodesmium* or cyanobacterial symbionts of large diatoms (another known diazotroph in these waters) were detected in all samples of the upper water column (0–150 m) collected in both February and May 2000, and displayed apparent spatial and diel variability (Fig. 1). The signal—detected by reverse transcription with polymerase chain reaction (RT-PCR)—at a depth of 25 m, indicating *nifH* biosynthesis, was highest at night (Fig. 1a), whereas the signal at greater depths was higher at noon than at midnight (Fig. 1b). Variation in transcript abundance during the day is a characteristic of nitrogenase expression pattern in unicellular and filamentous cyanobacteria, and is controlled by a circadian rhythm¹². Little or no *nifH* mRNA was detected at 150 m at any time of the day, which is consistent with the presence of relatively high nitrate concentrations, which selects against N₂ fixation (NO₃[−] concentration was 9.6 nM in the upper 100 m of the water column, compared with greater than 1,000 nM at 150 m; <http://hahana.soest.hawaii.edu>).

To identify which organisms in the 0.2–10-μm size class were expressing *nifH*, the *nifH* gene fragments that were amplified by RT-PCR were cloned and sequenced. Sequences phylogenetically related to cyanobacterial *nifH* genes were obtained from all depths (Fig. 2) and dominated the clone library (19 out of 27 clones). Eight clones clustered with sequences from proteobacteria, indicating that proteobacteria could also be fixing N₂, a system that could potentially be driven by bacterial, phototrophic metabolic pathways^{13,14}. Two different cyanobacterial *nifH* sequence groups were found (Fig. 2).

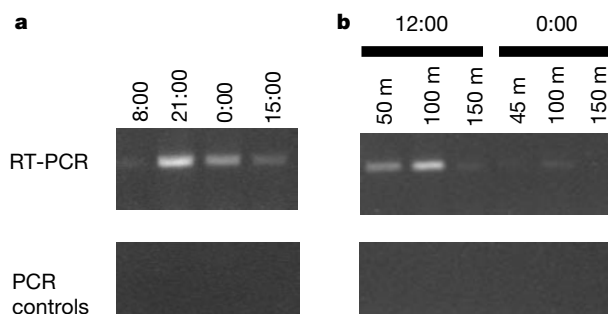


Figure 1 Expression of *nifH* genes in water samples collected at the Hawaii Ocean Time-series station ALOHA in May 2000 (HOT cruise 115). Upper gel on each panel shows amplified *nifH* fragment after RT-PCR; lower part shows the no-RT-PCR controls. No amplification was detected in samples treated with RNase. **a**, Amplification from water samples collected at 25 m at four different times of day. Highest levels of transcripts were detected at 21:00 and 0:00, with lower expression at 15:00 and 18:00. **b**, Amplification of *nifH* from water samples collected at 50, 100 and 150 m at 12:00 and 0:00.

One group of sequences (group A; Fig. 2), obtained from depths of 25, 50 and 100 m, were greater than 95% identical to cyanobacterial *nifH* sequences previously obtained by amplification of *nifH* from DNA collected at station ALOHA in 1996 and 1997 (ref. 10). The second group of sequences (group B) was found at a depth of 100 m in May and at various depths in February. The sequences recovered by RT-PCR within each group were between 93 and 99% identical to each other, which is more variability than can be accounted for by PCR error. Thus, the molecular data indicate that there are multiple strains or populations of unicellular, N₂-fixing marine cyanobacterial groups.

The cyanobacterial *nifH* genes amplified from the samples collected at station ALOHA were closely related to *nifH* genes of cultivated unicellular cyanobacteria, including the oceanic *Synechocystis* sp. strain WH 8501, previously isolated from the Atlantic Ocean (Figs 2 and 3c, d)^{15,16}. The genera of 3–10- μ m diameter, N₂-fixing unicellular cyanobacteria (for example *Synechocystis* sp. WH 8501, *Cyanothece* sp. ATCC 51142, and *Cyanothece* spp. Miami BG 43522 and 43511, which were previously called *Synechococcus*) have not previously been considered important components of the oceanic plankton, although phycoerythrin-containing cells of this size have been observed in oligotrophic waters at abundances much higher than we observed in July 2000 at station ALOHA^{17,18}.

As the evidence from the molecular sequences suggested the presence of cyanobacteria that are phylogenetically related to cultivated strains of 3–10- μ m diameter cyanobacteria, we examined water samples collected at station ALOHA in July 2000 by fluorescence microscopy, and attempted to cultivate N₂-fixing cyanobacteria. Phycoerythrin-containing cells, several micrometres in diameter, were observed by fluorescence microscopy in water samples collected from HOT cruise 117 at station ALOHA in July 2000 (Fig. 3). Different morphologies of cells, in particular different sizes of cells (Fig. 3), were observed, indicating that different strains or populations may be present in the water column, which is consistent with the observed diversity of *nifH* sequences (Fig. 2). Two strains of unicellular cyanobacteria (3–5 μ m in diameter) were cultivated on a medium containing no fixed N. The *nifH* gene sequences from these strains were 93% identical to cluster B sequences from station ALOHA transcripts (Fig. 2), and are very closely related to sequences from a previously cultivated isolate from the Atlantic, WH8501 (99–100% identical). The cultivated isolates fix N₂ at night in culture, which is consistent with the expression pattern observed at station ALOHA at a depth of 25 m (Fig. 1).

Water collected from a depth of 25 m in July 2000 was incubated with ¹⁵N₂ to determine if the size class containing marine unicellular cyanobacteria, but not *Trichodesmium* or *Rhizosolenia*–*Richelia*,

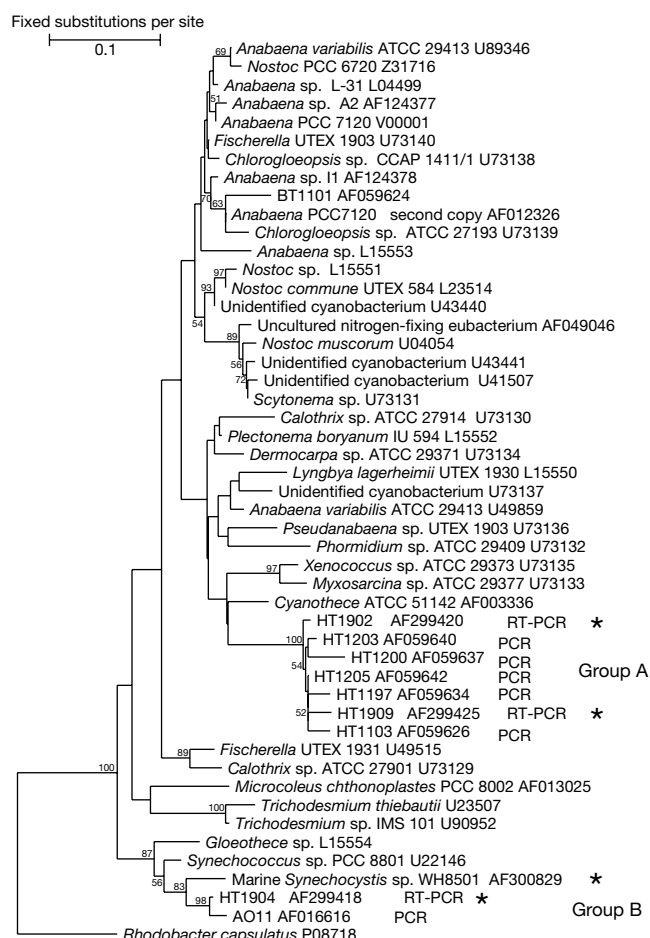


Figure 2 Phylogenetic tree showing the relationships among representative cyanobacterial *nifH* gene sequences. Transcripts detected at station ALOHA are designated RT-PCR. Nitrogenase (*nifH*) sequences amplified from station ALOHA are most closely related to *nifH* sequences from *Cyanothece*, *Myxosarcina*, *Gloeotheca* and *Synechococcus* (renamed *Cyanothece*). Two different clusters of sequences were detected in May 2000. Sequence type A was found at depths of 25, 50 and 100 m;

sequences of type B were detected in February and May 2000. The latter sequences are closely related to sequence AF016616, which was obtained from the equatorial North Atlantic Ocean. Sequences marked with an asterisk were determined as part of this study. The analysis was bootstrapped and bootstrap values greater than 50% (out of 100 replicates) are indicated at the respective nodes.

fixed $^{15}\text{N}_2$ under simulated *in situ* conditions. $^{15}\text{N}_2$ was fixed into particulate material in the 0.2–10- μm size fraction at rates of 10–16 $\mu\text{mol N l}^{-1} \text{ h}^{-1}$. Unicellular cyanobacterial cells 3–10 μm in diameter were present during the experiment at a concentration of $52 \pm 12 \text{ cells ml}^{-1}$ (s.d.; $n = 4$). This is equivalent to a cell-specific rate of N_2 fixation in the size fraction containing the natural populations of marine *Synechocystis* at station ALOHA that is similar to the rates reported for cultures of the morphologically similar *Cyanothece* sp. ATCC 51142 grown without combined N.

Although unicellular diazotrophs in the open ocean have received little attention, there is evidence that the N_2 -fixing unicellular cyanobacteria (that is, the 3–10 μm diameter morphology) may be widely distributed in marine and even freshwater environments at concentrations of up to 1,000 cells ml^{-1} (refs 17–20). The size and abundance of these diazotrophs indicate that they can equal or exceed the N_2 -fixation contribution of the known diazotrophs—*Trichodesmium* and *Richelia*. For example, *Trichodesmium* is found both as microscopic free filaments called trichomes (chains of approximately 100–150 cells) and as macroscopic aggregates composed of several hundred trichomes. The mean abundance of free trichomes at station ALOHA is $5 \times 10^4 \text{ m}^{-3}$ (ref. 21), equivalent to an average density of 5 cells ml^{-1} . Inclusion of the cells contained in the aggregate morphology would double or triple the mean *Trichodesmium* cell abundance. On the other hand, not all cells in a *Trichodesmium*

filament fix N_2 (ref. 22), and thus rates of N_2 fixation by *Trichodesmium* that are calculated from abundance data may be overestimated in some cases.

Concentrations of phycoerythrin-containing unicellular cyanobacteria cells at a diameter of 3–20 μm are present with a mean abundance of 10–50 cells ml^{-1} and can reach 1,000 cells ml^{-1} under bloom conditions¹⁸. For example, in April 1993, 3–20- μm diameter unicellular cyanobacteria were present at station ALOHA at an average abundance of 12 cells ml^{-1} in water samples from the surface to a depth of 187 m—there was slightly higher abundance in May¹⁸. In September, the unicellular cyanobacteria had an average abundance of 122 cells ml^{-1} from 0 to 150 m, with a maximum density of 928 cells ml^{-1} at 110 m (ref. 18). At these abundances, the biomass of the unicellular cyanobacteria could equal or exceed that of *Trichodesmium*. The bloom in September 1993, with an average of 122 cells ml^{-1} throughout a 150-m water column ($1.22 \times 10^6 \text{ cells m}^{-3}$; 150 m^3 in 150-m water column) may have been responsible for a N_2 -fixation rate of $92 \mu\text{mol m}^{-2} \text{ d}^{-1}$, assuming 10 h of night-time activity at an average rate of $0.5 \text{ fmol cell}^{-1} \text{ h}^{-1}$. This is the same order of magnitude as the averaged daily N_2 -fixation contribution of *Trichodesmium* (average of $137 \mu\text{mol m}^{-2} \text{ d}^{-1}$ (ref. 23)). However, owing to the well-documented temporal variations of N_2 -fixing and non- N_2 -fixing microorganisms at station ALOHA^{18,21,24}, determining the annual per cent contribution of N_2 fixation to new and export production will probably require a high-frequency observation programme over several years.

Our results indicate that N_2 -fixing cyanobacteria, in addition to the well-known diazotroph genera *Trichodesmium* and *Richelia*, are expressing *nifH* and fixing N_2 in the subtropical North Pacific Ocean. The results reported here point to a substantially new model for N_2 fixation in the oceans. The role of unicellular cyanobacteria in N_2 fixation in the open ocean is an unstudied phenomenon of uncertain ecological significance. Even if it is eventually found that these microorganisms do not contribute fixed N at rates equivalent to *Trichodesmium* (or other diazotrophs), it is probable that these unicellular microorganisms have a very different fate in the food chain with different implications for the fate of carbon in the water column. For example, the marine N_2 -fixing unicellular cyanobacteria may be more heavily consumed than *Trichodesmium*, which is toxic to some grazers²⁵. The documented temporal variations in the distributions and abundances of these populations, including aperiodic blooms, suggests that climate variations may induce shifts in the relative size distribution of microbial assemblages that fix N_2 . These new cyanobacterial diazotrophs may have long been, or are perhaps becoming, significant contributors to oceanic N_2 fixation, and deserve further scrutiny. The discovery that these microorganisms are present and actively expressing nitrogenase in the open ocean implies that conceptual models of the magnitude, timing and control of N_2 fixation in the ocean need to be re-evaluated²⁶. □

Methods

Sample collection

We collected water samples from the HOT station ALOHA ($22^\circ 45' \text{ N}$, $158^\circ 0' \text{ W}$) on HOT cruise 115. Samples were collected in 12-l polyvinylchloride bottles mounted on an aluminum rosette frame equipped with a SeaBird model 911 CTD and 24-place pylon. We collected samples from depths spanning the euphotic zone (25, 50, 100 and 150 m). Details of the HOT programme have been published elsewhere and are available at http://hahana.soest.hawaii.edu/hot/hot_jgofs.html.

Sample processing

Water samples were pre-filtered through plankton netting (pore size of 10 μm) to remove *Trichodesmium* and large phytoplankton, including diatoms, and then cells were collected on Millipore Durapore membrane filters (pore size of 0.22 μm). RNA was extracted from filters using the Qiagen RNeasy kit. RT-PCR was performed with the Promega RT-PCR Access protocol on RNA using degenerate primers for *nifH*²⁷. The nested PCR step was performed using a second inner set of degenerate *nifH* primers. The amplified products were separated on a 1.5% agarose gel and photographed with a Bio-Rad Gel Doc photodocumentation system.

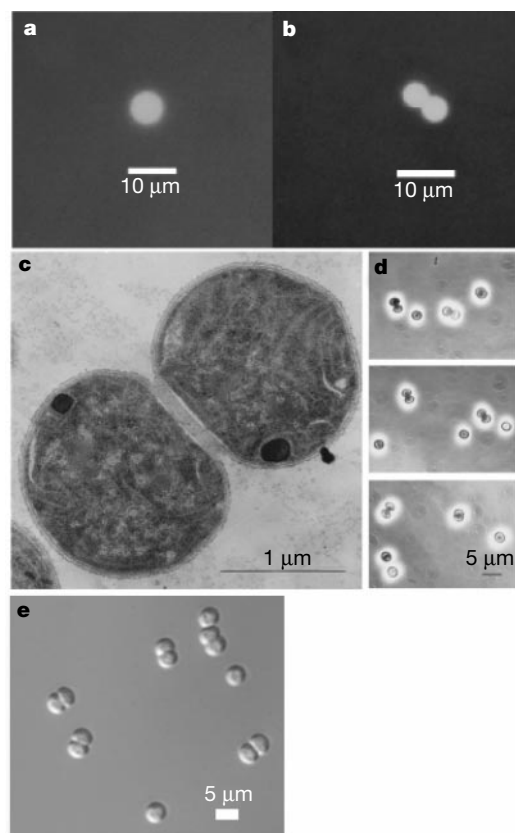


Figure 3 Photomicrographs of marine N_2 -fixing unicellular cyanobacteria. **a**, Fluorescence images of phycoerythrin-containing, 7- μm diameter unicellular cyanobacteria cells collected from 25 m depth at station ALOHA in July 2000. **b**, Fluorescence images of 3- μm diameter unicellular cyanobacteria cells that have recently divided, from the same sample as **a**. **c**, **d**, Electron (**c**) and light (**d**) micrographs of marine *Synechocystis* sp. WH 8501, a unicellular cyanobacterium isolated from tropical Atlantic Ocean waters (28° S , 43° W) that is phylogenetically related to group B unicellular cyanobacteria on the basis of *nifH* sequences (see Fig. 2). Cells are spherical and 2–3 μm in diameter. **e**, Unicellular cyanobacteria in enrichment in a fixed N-free medium, cultivated from surface water collected at station ALOHA in July 2000.

Cloning and sequencing

The amplified *nifH* fragments, approximately 359 base pairs in length, were cloned in Promega pGEM-T vector using the manufacturer's protocol. Recombinants containing the cloned insert were identified by restriction fragment analysis, and the fragment sequenced using an ABI 310 automated DNA sequencer. The translated sequences were aligned with Genetics Computer Group software, and the phylogenetic relationship determined by a distance method and neighbour joining using TREECON software²⁸. The analysis was bootstrapped 100 times.

Epifluorescence microscopy

Formalin-preserved water samples (10–20 ml) were filtered onto black polycarbonate Poretics filters (pore size 0.22 µm). An Olympus BX-60 epifluorescence microscope was used with the U-WMB filter set (excitation 450–480 nm; emission greater than 515 nm). Phycoerythrin-containing cells (fluorescing yellow–orange) were enumerated in 100 fields, and cells were measured using an ocular micrometer.

¹⁵N experiments

Water collected from a depth of 25 m at station ALOHA in July 2000 was transported to the marine laboratory at the Hawaii Institute of Marine Biology, where it was incubated in 2–4-l acid-washed polycarbonate bottles in running seawater, in a flume. The flume was covered with one layer of neutral-density screening to decrease light levels to roughly 50% surface irradiance (photosynthetically active radiation) approximately 600 µmol m⁻² s⁻¹. At each time point, samples were processed for ¹⁵N and microscopy, after pre-filtering through Nuclepore filters (pore size 10 µm) to remove the larger diazotrophs. Particle samples were collected by gentle pressure filtration (<10 pounds per square inch) through pre-combusted GF/F filters, which were analysed ashore by continuous-flow isotope ratio mass spectrometry using a Micromass Optima mass spectrometer that was interfaced with a CE-Elantech NA2500 elemental analyser. The analytical precision varies with sample size, but is about ± 0.2‰ (s.d. of replicate analyses) for samples containing 0.5–2 µmol of N.

Received 12 February; accepted 12 June 2001.

- Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry* **13**, 87–115 (1991).
- Howarth, R. W. & Marino, R. Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 2. Biogeochemical controls. *Limnol. Oceanogr.* **33**, 688–701 (1988).
- Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. *Nature* **387**, 272–275 (1997).
- Haug, G. H. et al. Glacial/interglacial variations in production and nitrogen fixation in the Cariaco Basin during the last 580 kyr. *Paleoceanography* **13**, 427–432 (1998).
- Paeerl, H. W. Physiological ecology and regulation of N₂ fixation in natural waters. *Adv. Microb. Ecol.* **8**, 305–344 (1990).
- Dugdale, R. C., Menzel, D. W. & Ryther, J. H. Nitrogen fixation in the Sargasso Sea. *Deep-Sea Res.* **7**, 298–300 (1961).
- Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochem. Cycles* **11**, 235–266 (1997).
- Michaels, A. F. et al. Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. *Biogeochemistry* **35**, 181–226 (1996).
- Lipschultz, F. & Owens, N. J. P. An assessment of nitrogen fixation as a source of nitrogen to the North Atlantic Ocean. *Biogeochemistry* **35**, 261–274 (1996).
- Zehr, J. P., Mellon, M. T. & Zani, S. New nitrogen fixing microorganisms detected in oligotrophic oceans by the amplification of nitrogenase (*nifH*) genes. *Appl. Environ. Microbiol.* **64**, 3444–3450 (1998).
- Zehr, J. P., Carpenter, E. J. & Villareal, T. A. New perspectives on nitrogen-fixing microorganisms in tropical and subtropical oceans. *Trends Microbiol.* **8**, 68–73 (2000).
- Chen, Y.-B., Dominik, B., Mellon, M. T. & Zehr, J. P. Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. Strain IMS 101. *J. Bacteriol.* **180**, 3598–3605 (1998).
- Beja, O. et al. Bacterial rhodospin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
- Kolber, Z. S., Van Dover, C. L., Niederman, R. A. & Falkowski, P. G. Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407**, 177–179 (2000).
- Reddy, K. J., Haskell, J. B., Sherman, D. M. & Sherman, L. A. Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus *Cyanothece*. *J. Bacteriol.* **175**, 1284–1292 (1993).
- Waterbury, J. B. & Rippka, R. in *Bergey's Manual of Systematic Bacteriology* Vol. 3 (ed. Staley, J. T.) 1728–1746 (Williams & Wilkins, Baltimore, 1989).
- Neveux, J., Lantoiné, F., Vaulot, D., Marie, D. & Blanchot, J. Phycoerythrins in the southern tropical and equatorial Pacific Ocean: evidence for new cyanobacterial types. *J. Geophys. Res.* **104**, 3311–3321 (1999).
- Campbell, L., Liu, H., Nolla, H. A. & Vaulot, D. Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at station ALOHA during the 1991–1994 ENSO event. *Deep-Sea Res.* **144**, 167–192 (1997).
- Brass, S. et al. Utilization of light for nitrogen fixation by a new *Synechocystis* strain is extended by its low photosynthetic efficiency. *Appl. Environ. Microbiol.* **60**, 2575–2583 (1994).
- Wasmund, N., Voss, M. & Lochte, K. Evidence of nitrogen fixation by non-heterocystous cyanobacteria in the Baltic Sea and re-calculation of a budget of nitrogen fixation. *Mar. Ecol. Progr. Ser.* **214**, 1–14 (2001).
- Letelier, R. M. & Karl, D. M. Role of *Trichodesmium* spp. in the productivity of the subtropical North Pacific Ocean. *Mar. Ecol. Progr. Ser.* **133**, 263–273 (1996).
- Lin, S., Henze, S. & Carpenter, E. J. Whole-cell immunolocalization of nitrogenase in marine diazotrophic cyanobacteria, *Trichodesmium* spp. *Appl. Environ. Microbiol.* **64**, 3052–3058 (1998).

- Karl, D. et al. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
- Letelier, R. M. et al. Temporal variability of phytoplankton community structure based on pigment analysis. *Limnol. Oceanogr.* **38**, 1420–1437 (1993).
- Hawser, S. P., O'Neil, J. M., Roman, M. R. & Codd, G. A. Toxicity of blooms of the cyanobacterium *Trichodesmium* to zooplankton. *J. Appl. Phys.* **4**, 79–86 (1992).
- Karl, D. Comment: a new source of 'new' nitrogen in the sea. *Trends Microbiol.* **8**, 301 (2000).
- Zani, S., Mellon, M. T., Collier, J. L. & Zehr, J. P. Expression of *nifH* genes in natural microbial assemblages in Lake George, NY detected with RT-PCR. *Appl. Environ. Microbiol.* **66**, 3119–3124 (2000).
- Van de Peer, Y. & De Wachter, R. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Applic. Biosci.* **10**, 569–570 (1994).

Acknowledgements

We thank the crews and participants of the HOT programme, especially L. Tupas and J. Dore for field support. We also thank L. Campbell for providing flow cytometer data. This work was supported by NSF Division of Ocean Sciences grants to J.P.Z., J.P.M. and D.M.K.

Correspondence and requests for materials should be addressed to J.P.Z. (e-mail: zehrj@cats.ucsc.edu).

Density-dependent mortality in an oceanic copepod population

M. D. Ohman*† & H.-J. Hirche‡

* Station Zoologique, 06230 Villefranche-sur-Mer, France

‡ Alfred Wegener Institute for Polar and Marine Research, D-27568 Bremerhaven, Germany

Planktonic copepods are primary consumers in the ocean and are perhaps the most numerous metazoans on earth. Secondary production by these zooplankton supports most food webs of the open sea, directly affecting pelagic fish populations and the biological pump of carbon into the deep ocean. Models of marine ecosystems are quite sensitive to the formulation of the term for zooplankton mortality^{1–4}, although there are few data available to constrain mortality rates in such models. Here we present the first evidence for nonlinear, density-dependent mortality rates of open-ocean zooplankton. A high-frequency time series reveals that per capita mortality rates of eggs of *Calanus finmarchicus* Gunnerus are a function of the abundance of adult females and juveniles. The temporal dynamics of zooplankton populations can be influenced as much by time-dependent mortality rates as by variations in 'bottom up' forcing. The functional form and rates chosen for zooplankton mortality in ecosystem models can alter the balance of pelagic ecosystems^{1–3}, modify elemental fluxes into the ocean's interior⁵, and modulate interannual variability in pelagic ecosystems⁶.

The high-frequency (sampling interval 1–2 d, sustained for 80 d) time series of *Calanus finmarchicus* in the central Norwegian Sea conducted as part of the TASC (Trans-Atlantic Study of *Calanus finmarchicus*) programme shows a springtime emergence of juvenile (C5) and adult female copepods in late March (small hump at front left of Fig. 1). These individuals originate from the overwintering generation in deep water⁷. The onset of reproductive maturity and egg production occur at least 40–50 d before the phytoplankton bloom^{8,9}, contrary to the assumptions in many ecosystem models, although per capita rates of egg production increase at the time of

† Present address: Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0218, USA.

the bloom⁹. Grazing rates by *C. finmarchicus* are low before the bloom^{10,11}, suggesting that stored wax esters are used during gonad maturation and oogenesis with nutritional supplements obtained from microplankton grazed from the water column. The recruitment rate (R_e ; number of eggs produced per m^2 of sea surface per d) is initially quite high, but for the first 20 d of the series virtually no eggs survive to the first larval stage (the nauplius). Subsequently, nauplius larvae appear and the population increases as a developing spring generation, which is apparent as a dark diagonal band across Fig. 1.

Stage-specific mortality rates from this series were estimated using inverse methods^{12,13} (see Fig. 2 legend and Methods for details). The high rates of instantaneous egg mortality (median = $1.76 d^{-1}$, equivalent to $82.8\% d^{-1}$) indicate that on average only 2.7% of eggs in this population survive to the first larval stage. The median rate of egg death is balanced by the median instantaneous birth rate ($1.71 d^{-1}$), but this seasonally averaged equilibrium masks the important temporal dynamics in the population. Rates of instantaneous egg mortality vary considerably with time. They peak on day 94–95, generally decline until day 129, then increase slowly after this period (Fig. 2a). Temporal variations in rates of egg mortality are greater than variations in birth rate (Fig. 2b).

Such losses cannot be attributed to advection alone. The sustained lack of nauplii for nearly 20 d—when eggs hatch in 2.16 d at

the ambient temperature of $6.4^\circ C$ (ref. 14)—implies that eggs perish consistently in upstream waters as well as at Ocean Station M. The residence time of a parcel of water equivalent to the dimensions of a 20-km *Calanus* patch¹⁵ can be estimated at 5 d from the mean springtime current speeds in the upper 100 m at Ocean Station M¹⁶. This residence time gives an average fractional daily loss due to advection of $20\% d^{-1}$, while the average rate of egg mortality is four times greater ($83\% d^{-1}$). A third line of evidence indicating that we have sampled the same planktonic population is the agreement between seasonally averaged rates of birth and mortality.

The highest rates of egg mortality occur at a time of very low diatom concentrations^{10,11} and thus are not consistent with deleterious effects of diatom metabolites¹⁷. Using fucoxanthin pigments as a measure of the biomass of diatoms¹¹, rates of *C. finmarchicus* egg mortality were not correlated with the concentration of diatoms in the water column ($P > 0.10$, Spearman's rank correlation, $n = 20$) or with the rate of ingestion of diatoms by *C. finmarchicus* adult females ($P > 0.10$, $n = 19$). The highest rates of egg mortality occur when adult males are most abundant¹⁸, when the adult male:female sex ratio is high (approximately 1:2)⁹, when immature females are relatively rare⁹, and when adult females all have full seminal receptacles (H.-J.H., unpublished observations). Such observations suggest that fertilization success is unlikely to account for high rates

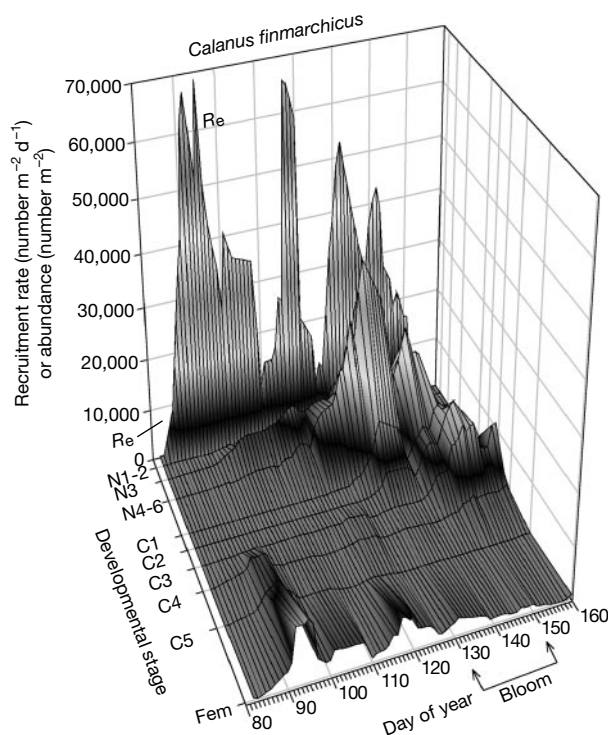


Figure 1 Spring population growth of the copepod *Calanus finmarchicus* at ocean station M, in the Norwegian Sea (centred on $66^\circ N$, $2^\circ E$). Bloom indicates the period of the phytoplankton bloom (maximum concentration 3 ng chlorophyll *a* per ml), which coincided with the onset of thermal stratification (see ref. 10). Pre-bloom concentrations were <0.5 ng chlorophyll *a* per ml and post-bloom averaged 1.5 ng chlorophyll *a* per ml (ref. 10). The population was sampled with a $53\text{-}\mu m$ mesh WP2 net, towed vertically from 100 to 0 m, every 1–2 d from 22 March to 9 June 1997 (day of year 81–160). Daily rates of egg production were determined from 50 freshly collected individual females incubated aboard ship⁹, multiplied by the abundance of adult females to obtain the daily rate of egg recruitment (R_e ; eggs $m^{-2} d^{-1}$). Other values are abundances (number m^{-2}) of the following developmental stages: nauplius stages 1 and 2 combined (N1–2), N3, nauplius 4–6 combined (N4–6), copepodid stages C1 through to C5, and adult females (for details see refs 9 and 18). The raw data (see ref. 18) were smoothed with a four-point running mean and interpolated to a rectangular grid for plotting.

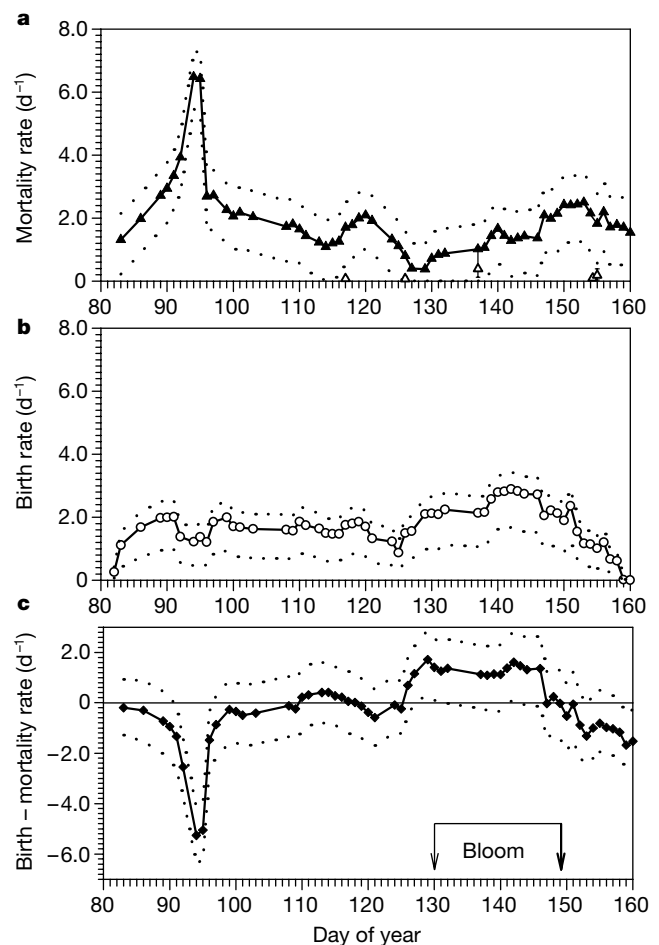


Figure 2 Instantaneous rates of egg mortality (a) (m_e ; d^{-1}), birth (b) (b ; d^{-1}) and their difference (c) ($b - m_e$) at ocean station M, with estimated 95% confidence intervals (dotted lines). Filled triangles in panel a show total egg mortality; open triangles show mortality attributable to egg-hatching success ($\bar{x} \pm 95\%$). Birth rates were obtained from $b = \ln(1 + R_e/2N_f)$, where R_e is rate of egg recruitment (eggs $m^{-2} d^{-1}$) and N_f is the number of adult females, assuming that the egg sex ratio was 1:1. See Methods for the calculation of embryonic mortality.

of egg mortality. If egg mortality were caused by a deficiency in micronutrients that are essential for embryonic development, this deficiency would be expected to diminish when food concentrations increase. Instead, rates of egg mortality were independent of the biomass of ciliate prey, or phytoplankton, and of the two prey types combined ($P > 0.10$, data from ref. 10), although we cannot rule out micronutrient deficiency definitively. Experiments on egg-hatching success on five occasions showed that mortality attributable to the production of nonviable eggs was a small fraction (average 10.2%) of total egg mortality (Fig. 2a).

The temporal decline in egg mortality rate (Fig. 2a) parallels the decline in abundance of adult female and late-stage juvenile copepods (Fig. 1). We found variations in the rate of instantaneous egg mortality to be directly proportional to the abundance of adult female and juvenile *C. finmarchicus* (Fig. 3, $P < 0.001$). The instantaneous mortality rate represents the probability of mortality; where this probability changes as a function of population density, it introduces a nonlinear term into equations of population dynamics³. In contrast to rates of egg mortality, birth rates were independent of the abundance of adult female and juvenile *C. finmarchicus* ($P > 0.05$).

The most parsimonious explanation for the density-dependence of egg mortality seems to be egg cannibalism. Experiments have established that *C. finmarchicus* adult females and stage C5 juveniles ingest eggs of conspecifics (J. A. Runge, personal communication) and similarly sized microzooplankton^{10,19} with high clearance rates. Egg predation would be particularly important during the pre-bloom period when phytoplankton concentrations are low. Application of experimentally measured clearance rates to adult females and stage C5 juveniles of *C. finmarchicus* show that if the copepods encounter patches of eggs in high concentrations in thin layers, or if encounter rates between *Calanus* and eggs are increased by ambient turbulence²⁰, experimentally measured rates of cannibalism are typically sufficient to account for daily rates of egg loss from the water column. However, such calculations depend on the concentrations of eggs encountered by copepods in the water column at the microscale, a problem that merits future research. The residuals from the regression in Fig. 3 are not consistently negative during the

spring bloom, as would be expected if the copepods were switching (see ref. 21) to alternative prey during the bloom. Although copepod egg cannibalism has been shown in laboratory containers and cannibalism has been inferred for natural copepod populations in semi-closed coastal environments^{22–24} and shallow continental shelf localities^{25,26}, this seems to be the first such suggestion for a zooplankton population of the open sea.

Why does reproduction persist early in the year if egg survivorship is so low? There is substantial interannual variability in the population size of *C. finmarchicus* in the North Atlantic²⁷. If cannibalism were the dominant mechanism explaining early egg mortality, year-to-year differences in the abundance of *Calanus* females would result in year-to-year differences in numbers of eggs surviving. An occasional year with low egg mortality would be sufficient to maintain the trait of early reproduction in the population.

The difference between birth and mortality rates ($b - m_e$) becomes approximately equal to 0 at day 100 (Fig. 2c), at which time the population begins increasing (Fig. 1). Although b subsequently increases at the time of the spring bloom (after day 130, Fig. 2b), the increase is preceded by an 8–10-d decline in egg mortality (Fig. 2a), which seems to initiate increased population growth shortly before the bloom (Fig. 2c). Thus the onset of net population growth seems to be affected at least as much, if not more, by changes in the rate of egg death as by variations in birth rate.

Traditional models of marine ecosystems emphasize the physical and biological processes influencing nutrient supply, dynamics of the phytoplankton spring bloom, and zooplankton grazing, on the assumption that this bottom-up perspective will predict the fecundity and production of marine zooplankton populations. Rates of zooplankton loss are often treated as constant and linear^{28,29}. The present results show that temporal variations of mortality rates, perhaps more than fecundity responses to a spring bloom, are of central importance in explaining the temporal variability of the dominant oceanic zooplankton in the North Atlantic. Mortality acting in a density-dependent manner can introduce self-limitation of some zooplankton populations of the open sea. Predictions of long-term variability in pelagic ecosystems cannot rely on only changes in ocean mixing and circulation^{27,28,30} but must also consider internal nonlinearities operating within pelagic populations. □

Methods

Embryonic mortality rates as a function of time

The calculation of the rates of embryonic mortality as a function of time ($m_e(t)$) (used to generate Fig. 2) were derived from a new solution to equations (5) and (11) of ref. 13:

$$m_e(t) = \frac{-\ln\left(\frac{N_{n1}(t)m_{n1}(t)}{R_e(t - a_e(t))(1 - e^{-m_{n1}(t)a_{n1}(t)})}\right)}{a_e(t)}$$

where $N_{n1}(t)$ is the abundance of the combined nauplius stage 1–2 at time t ; $m_{n1}(t)$ is the mortality rate of nauplius stage 1–2 at time t ; $R_e(t - a_e(t))$ is the rate of egg recruitment at time $t - a_e(t)$; and $a_e(t)$ is the embryonic duration at time t . This solution assumes that temperature-dependent embryonic duration changes slowly with time (that is, $da_e/dt \approx 0$), as confirmed from field measurements of temperature, and that R_e and m_{n1} are constant for the duration of the nauplius stages N1–2, $a_{n1}(t)$. The value for $m_{n1}(t)$ was obtained using Wood's population surface method¹². The durations for eggs and naupliar stages N1–2 were obtained from Bełehrádek functions¹⁴, using the average temperature in the upper 100 m of the water column. As all of these developmental stages are before feeding, their rate of development is controlled by ambient temperature.

Confidence intervals were estimated using 10,000 iterations of a parametric bootstrap procedure assuming that the error in N_{n1} was normally distributed on a log scale with a coefficient of variation of 70%; that other variables had normally distributed errors with coefficients of variation of $R_e = 70\%$, $m_{n1} = 50\%$, and $R_e/N_e = 40\%$, respectively; and that a_e and a_{n1} were known parameters.

Mortality at egg hatching was determined by counting the number of nauplii hatched from lots of 30–60 eggs over an interval of 3 d ($n = 6$ to 12 replicates per treatment), and expressed as an instantaneous coefficient ($-\ln(L_x)/a_e$), where L_x is the proportion of eggs surviving and a_e is the temperature-dependent embryonic duration. Egg-hatching data were supplied by U. Klenke.

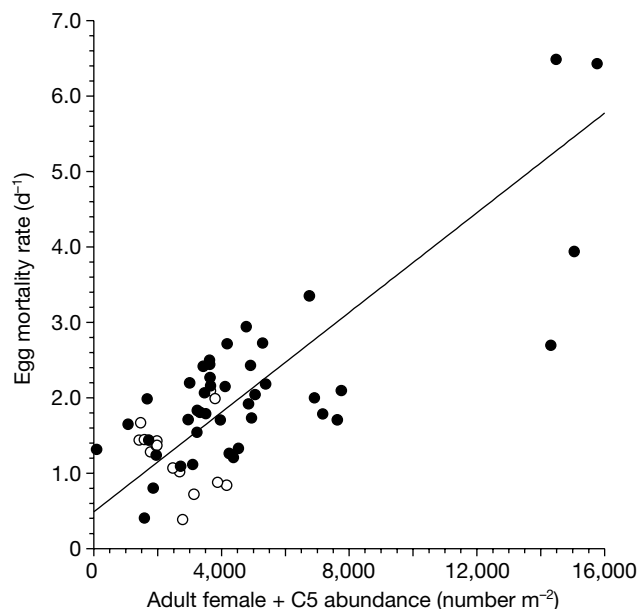


Figure 3 Dependence of embryonic mortality rates on the abundance of *Calanus finmarchicus* adult females and juvenile stage C5 at ocean station M. Filled symbols indicate the period before or after the phytoplankton bloom, open symbols the bloom period. The fitted line illustrates a model II functional regression, assuming error in both y and x values: $y = 0.00033x + 0.488$ ($r^2 = 0.608$, $P < 0.0001$).

Received 4 January; accepted 20 June 2001.

1. Steele, J. H. & Henderson, E. W. The role of predation in plankton models. *J. Plank. Res.* **14**, 157–172 (1992).
2. Fasham, M. J. R. Variation in the seasonal cycle of biological production in subarctic oceans: a model sensitivity analysis. *Deep-Sea Res.* **42**, 1111–1149 (1995).
3. Edwards, A. M. & Yool, A. The role of higher predation in plankton population models. *J. Plank. Res.* **22**, 1085–1112 (2000).
4. Carlotti, F., Giske, J. & Werner, F. *Zooplankton Methodology Manual* (eds Harris, R. P., Wiebe, P. H., Lenz, J., Skjoldal, H. R. & Huntley, M.) 571–667 (Academic, San Diego, 2000).
5. Fasham, M. J. R. *The Global Carbon Cycle* (ed. Heinemann, M.) 457–504 (Springer, New York, 1993).
6. Li, M., Gargett, A. & Denman, K. What determines seasonal and interannual variability of phytoplankton and zooplankton in strongly estuarine systems? Application to the semi-enclosed estuary of Strait of Georgia and Juan de Fuca Strait. *Estuar. Coast. Shelf Sci.* **50**, 467–488 (2000).
7. Heath, M. R. The ascent migration of *Calanus finmarchicus* from overwintering depths in the Faroe–Shetland Channel. *Fish. Oceanogr.* (Suppl. 1) **8**, 84–99 (1999).
8. Richardson, K., Jonasdottir, S. H., Hay, S. J. & Christoffersen, A. *Calanus finmarchicus* egg production and food availability in the Faroe–Shetland channel and northern North Sea: October–March. *Fish. Oceanogr.* (Suppl. 1) **8**, 153–162 (1999).
9. Niehoff, B. *et al.* A high frequency time series at Weathership M, Norwegian Sea, during the 1997 spring bloom: the reproductive biology of *Calanus finmarchicus*. *Mar. Ecol. Prog. Ser.* **176**, 81–91 (1999).
10. Irigoien, X. *et al.* A high frequency time series at weathership M, Norwegian Sea, during the 1997 spring bloom: feeding of adult female *Calanus finmarchicus*. *Mar. Ecol. Prog. Ser.* **172**, 127–137 (1998).
11. Meyer-Harms, B., Irigoien, X., Head, R. & Harris, R. Selective feeding on natural phytoplankton by *Calanus finmarchicus* before, during, and after the 1997 spring bloom in the Norwegian sea. *Limnol. Oceanogr.* **44**, 154–165 (1999).
12. Wood, S. N. Obtaining birth and mortality patterns from structured population trajectories. *Ecol. Monogr.* **64**, 23–44 (1994).
13. Aksnes, D. L., Miller, C. B., Ohman, M. D. & Wood, S. N. Estimation techniques used in studies of copepod population dynamics—a review of underlying assumptions. *Sarsia* **82**, 279–296 (1997).
14. Campbell, R. G., Wagner, M. M., Teegarden, G. J., Boudreau, C. A. & Durbin, E. G. Growth and development rates of the copepod *Calanus finmarchicus* reared in the laboratory. *Mar. Ecol. Prog. Ser.* (in the press).
15. Solow, A. R. & Steele, J. H. Scales of plankton patchiness: biomass versus demography. *J. Plank. Res.* **17**, 1669–1677 (1995).
16. Hainbucher, D. & Backhaus, J. O. Circulation of the eastern North Atlantic and north-west European continental shelf—a hydrodynamic modelling study. *Fish. Oceanogr.* (Suppl. 1) **8**, 1–12 (1999).
17. Miralto, A. *et al.* The insidious effect of diatoms on copepod reproduction. *Nature* **402**, 173–176 (1999).
18. Hirche, H.-J., Brey, T. & Niehoff, B. A high frequency time series at Weathership M, Norwegian Sea: population dynamics of *Calanus finmarchicus*. *Mar. Ecol. Prog. Ser.* (in the press).
19. Ohman, M. D. & Runge, J. A. Sustained fecundity when phytoplankton resources are in short supply: omnivory by *Calanus finmarchicus* in the Gulf of St. Lawrence. *Limnol. Oceanogr.* **39**, 21–36 (1994).
20. Rothschild, B. J. & Osborn, T. R. Small-scale turbulence and plankton contact rates. *J. Plank. Res.* **10**, 465–474 (1988).
21. Landry, M. R. Switching between herbivory and carnivory by the planktonic marine copepod *Calanus pacificus*. *Mar. Biol.* **65**, 77–82 (1981).
22. Landry, M. R. Population dynamics and production of a planktonic marine copepod, *Acartia clausii*, in a small temperate lagoon on an Juan Island, Washington. *Int. Rev. Ges. Hydrobiol.* **63**, 77–119 (1978).
23. Uye, S. I. & Liang, D. Copepods attain high abundance, biomass and production in the absence of large predators but suffer cannibalistic loss. *J. Mar. Sys.* **15**, 495–501 (1998).
24. Peterson, W. T. & Kimmerer, W. J. Processes controlling recruitment of the marine calanoid copepod *Temora longicornis* in Long Island Sound: egg production, egg mortality, and cohort survival rates. *Limnol. Oceanogr.* **39**, 1594–1605 (1994).
25. Daan, R., Gonzales, S. R. & Klein Breteler, W. C. M. Cannibalism in omnivorous calanoid copepods. *Mar. Ecol. Prog. Ser.* **47**, 45–54 (1989).
26. Ohman, M. D., Durbin, E. G. & Runge, J. A. Density-dependence of instantaneous mortality rates of *Calanus finmarchicus* on Georges Bank. *EOS Trans. Am. Geophys. Union* **79**, OS155 (1998).
27. Planque, B. & Taylor, A. H. Long-term changes in zooplankton and the climate of the North Atlantic. *ICES J. Mar. Sci.* **55**, 644–654 (1998).
28. Lynch, D. R., Gentleman, W. C., McGillicuddy, D. J. Jr & Davis, C. S. Biological/physical simulations of *Calanus finmarchicus* population dynamics in the Gulf of Maine. *Mar. Ecol. Prog. Ser.* **169**, 189–210 (1998).
29. Cowan, R. K., Lwiza, K. M. M., Sponaugle, S., Paris, C. B. & Olson, D. B. Connectivity of marine populations: open or closed? *Science* **287**, 857–859 (2000).
30. Heath, M. R. *et al.* Climate fluctuations and the spring invasion of the North Sea by *Calanus finmarchicus*. *Fish. Oceanogr.* (Suppl. 1) **8**, 163–176 (1999).

Acknowledgements

We thank the late M. M. Mullin for his scientific insights, the captain and the crew as well as the scientists (X. Irigoien, U. Klenke, R. Head) on the vessel *Polarfront* for their support, and the Institute for Marine Research (Bergen, Norway), which provided logistical help. B. Niehoff provided egg-production rates, S. Jaklin and E. Mizdalski helped with analysing the samples and A. De Robertis generated bootstrap confidence intervals. This work was supported by funding from the European Commission through the TASC project and by the National Science Foundation and the National Oceanic and Atmospheric Administration through US GLOBEC (Global Ocean Ecosystem Dynamics).

Correspondence and requests for materials should be addressed to M.D.O. (e-mail: mohman@ucsd.edu).

Erythropoietin-mediated neuroprotection involves cross-talk between Jak2 and NF- κ B signalling cascades

Murat Digicaylioglu & Stuart A. Lipton

Center for Neuroscience and Aging Research, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA
Cerebrovascular and Neuroscience Research Institute, Brigham and Women's Hospital, Program in Neuroscience, Harvard Medical School, Boston, Massachusetts 02115, USA

Erythropoietin, a kidney cytokine regulating haematopoiesis (the production of blood cells), is also produced in the brain after oxidative or nitrosative stress^{1,2}. The transcription factor hypoxia-inducible factor-1 (HIF-1) upregulates EPO following hypoxic stimuli^{3,4}. Here we show that preconditioning with EPO protects neurons in models of ischaemic and degenerative damage due to excitotoxins^{4,5} and consequent generation of free radicals, including nitric oxide (NO). Activation of neuronal EPO receptors

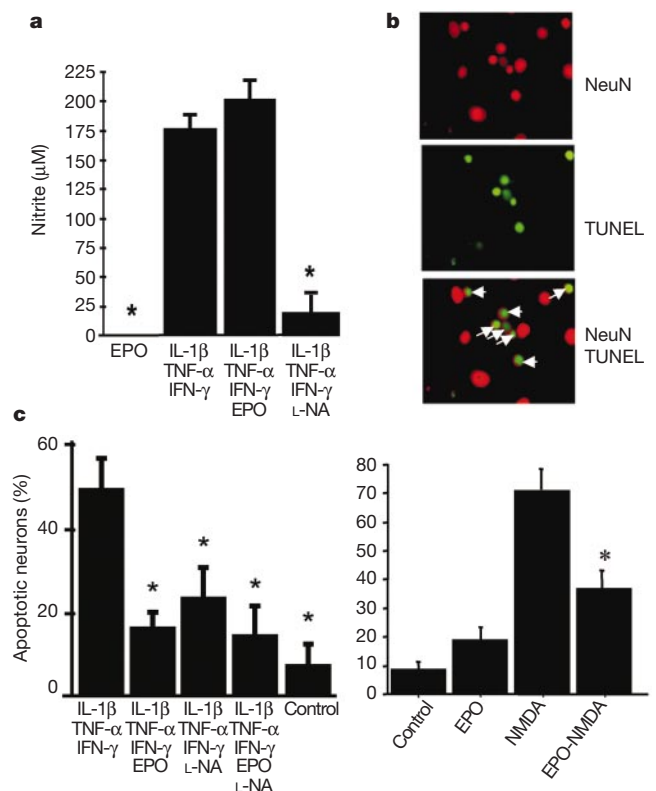


Figure 1 Neuroprotective effect of EPOR activation on cerebrocortical neurons.

a, Incubation of rat cerebrocortical cultures for 6 h with cytokines (IFN- γ , 500 U ml⁻¹; TNF- α , 200 U ml⁻¹; IL-1 β , 5 ng ml⁻¹) increased NO, as reflected by nitrite concentration. EPO (5 U ml⁻¹) did not affect nitrites, whereas L-nitroarginine (L-NA; 1 mM) dramatically reduced nitrite production; asterisk, $P < 0.001$ by analysis of variance (ANOVA).

b, Apoptotic neurons (arrows) identified by labelling with both anti-NeuN (red) and TUNEL (green). **c**, Neuronal apoptosis increased after NMDA (300 μ M) or cytokine-induced NO production, but decreased with EPO. Pre-incubation with EPO (5 U ml⁻¹, 3 h) dramatically decreased the number of apoptotic neurons (asterisk, $P < 0.001$). L-nitroarginine (1 mM) administered with cytokines also decreased the number of apoptotic neurons.

(EPORs) prevents apoptosis induced by NMDA (N-methyl-D-aspartate) or NO by triggering cross-talk between the signalling pathways of Janus kinase-2 (Jak2) and nuclear factor- κ B (NF- κ B). We show that EPOR-mediated activation of Jak2 leads to phosphorylation of the inhibitor of NF- κ B (I κ B), subsequent nuclear translocation of the transcription factor NF- κ B, and NF- κ B-dependent transcription of neuroprotective genes. Transfection of cerebrocortical neurons with a dominant interfering form of Jak2

or an I κ B α super-repressor blocks EPO-mediated prevention of neuronal apoptosis. Thus neuronal EPORs activate a neuroprotective pathway that is distinct from previously well characterized Jak and NF- κ B functions. Moreover, this EPO effect may underlie neuroprotection mediated by hypoxic-ischaemic preconditioning.

We initially characterized expression of EPOR protein on rat brain neurons by immunohistochemistry with a specific monoclonal antibody (see Supplementary Information)⁶. Next, we inves-

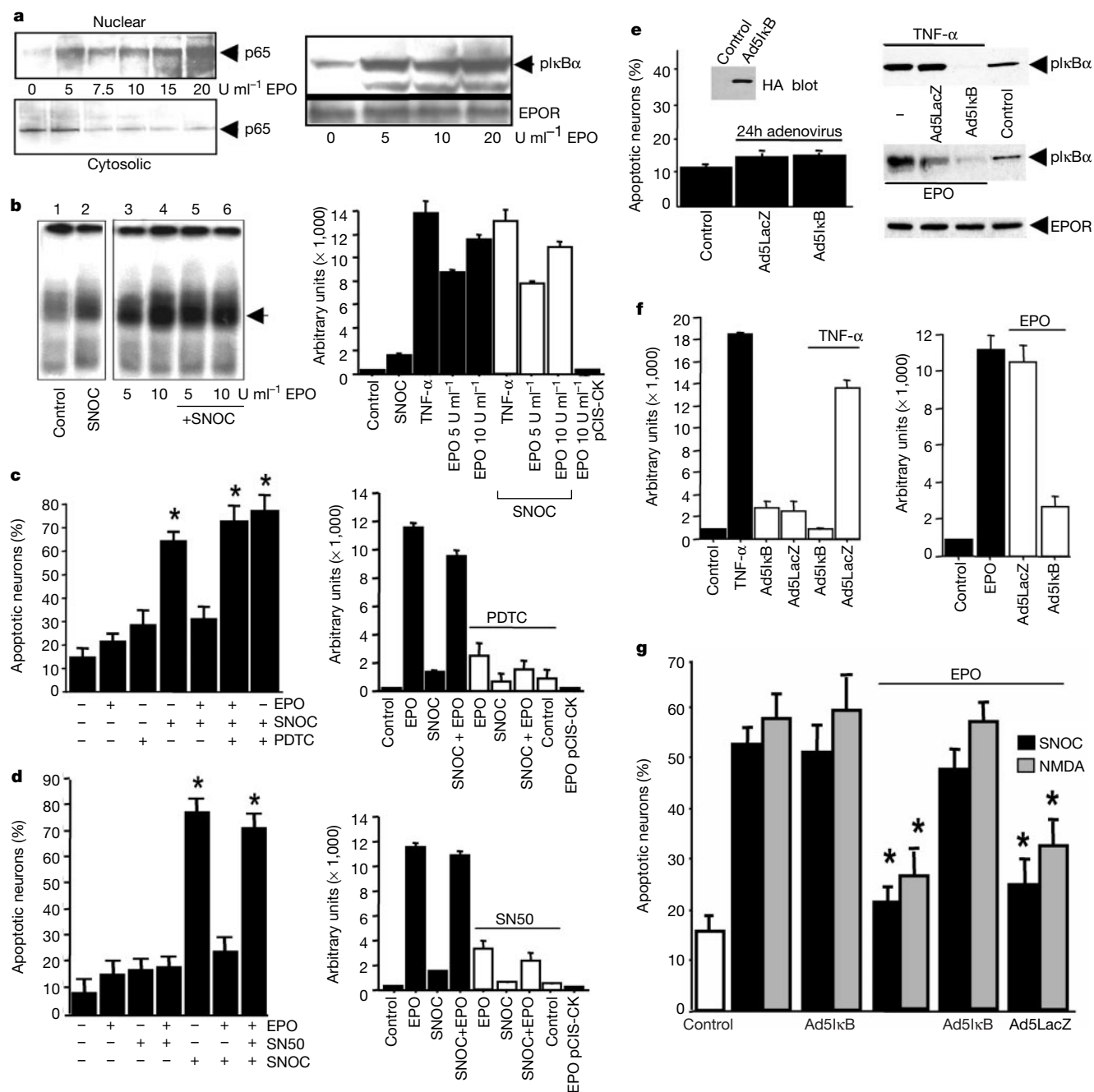


Figure 2 EPO-induced NF- κ B activation. **a**, EPO-induced nuclear accumulation and cytoplasmic depletion of NF- κ B (p65 subunit) in cerebrocortical cultures by immunoblot (left); increased cytoplasmic phosphorylated I κ B (pI κ B α) detected with anti-phospho-Ser^{32,36}I κ B α (right). **b**, EPO-induced DNA-binding activity of NF- κ B (arrow, left), and *trans*-activational activity of NF- κ B in reporter gene assays (right). SNOC, S-nitrosocysteine. **c**, Neuronal apoptosis induced by SNOC (200 μ M) was prevented by pre-incubation in EPO (5 U ml⁻¹). PDTC (5 μ M) blocked EPO-induced neuroprotection and NF- κ B *trans*-activational activity (asterisk, $P < 0.001$). **d**, Inhibition of NF- κ B *trans*-activational activity (right) by SN50 prevented EPO-mediated protection from SNOC-

induced neuronal apoptosis (left). **e**, Apoptosis was unaffected by Ad5I κ B or Ad5LacZ (left). Immunoblot confirmed expression of I κ B super-repressor tagged with haemagglutinin (HA) (inset). A tenfold infection of Ad5I κ B, but not of Ad5LacZ, inhibited EPO- and TNF- α -induced phosphorylation of I κ B α (detected with anti-phospho-Ser^{32,36}, right). **f**, Ad5I κ B abrogated increases in NF- κ B *trans*-activational activity observed with TNF- α (left) or EPO (right). **g**, Blocking NF- κ B activation with Ad5I κ B abolished the EPO neuroprotective effect (asterisk, $P < 0.0001$). Second and fourth sets of bars were obtained in the absence of adenoviral infection.

regulated intracellular signalling pathways underlying EPO-induced neuroprotection in cerebrocortical neurons in culture. In non-neuronal cells, EPO induces tyrosine phosphorylation of EPORs and its associated kinase, Jak2 (ref. 7), and we found similar results in neurons. In the brain, EPO increases in response to oxidative or nitrosative stress^{3–5}. NO can mediate this stress and contribute to neuronal death triggered by ischaemia, inflammation and neurodegenerative diseases. In these disorders, excitotoxic (glutamate) overactivation of NMDA receptors on neurons leads to excessive production of NO via Ca^{2+} stimulation of neuronal NO synthase (NOS)^{8,9}. Additionally, pro-inflammatory cytokines stimulate inducible NOS in astrocytes leading to NO production and neuronal damage. NO, after reaction with superoxide (O_2^-) to form peroxynitrite (ONOO^-), induces neuronal apoptosis if the insult is mild, or necrosis if it is more intense¹⁰. To investigate possible neuroprotective effects of EPO from endogenous NO, we incubated cerebrocortical cultures in cytokines. Stimulation by cytokines increased nitrites in the medium, reflecting increased NO, to $175 \pm 12 \mu\text{M}$ from a control of $11 \pm 4 \mu\text{M}$ (mean $\pm$ s.e.m., $n = 9$). EPO did not prevent this cytokine-induced increase; however, L-nitroarginine,

a NOS antagonist, reduced nitrite production to $11 \pm 3 \mu\text{M}$ (Fig. 1a). To assess the ability of EPO to block NMDA- or NO-induced neuronal apoptosis, we identified apoptotic neurons by double immunofluorescence with anti-NeuN (neuronal nuclei protein) to specifically label neurons, and with TUNEL (terminal-deoxynucleotidyl-transferase-mediated dUTP nick-end labelling) plus nuclear morphology to detect apoptosis (Fig. 1b). Pre-incubation with EPO significantly reduced neuronal apoptosis despite the cytokine-induced NO insult (Fig. 1c, left). Under these conditions, apoptosis was related to NO because inhibition of NOS with L-nitroarginine prevented neuronal death. L-nitroarginine and EPO did not have additive effects. Pre-incubation with EPO also prevented NMDA-induced neuronal apoptosis (Fig. 1c, right). As a control, heat-inactivated EPO did not enhance neuronal survival, nor did a variety of other proteins including bovine serum albumin (data not shown). Taken together, these results suggest that EPO protects from NMDA- and NO-related neuronal apoptosis but does not affect NO levels.

Next we characterized the intracellular signalling cascade initiated by EPO–EPOR interaction that leads to neuroprotection.

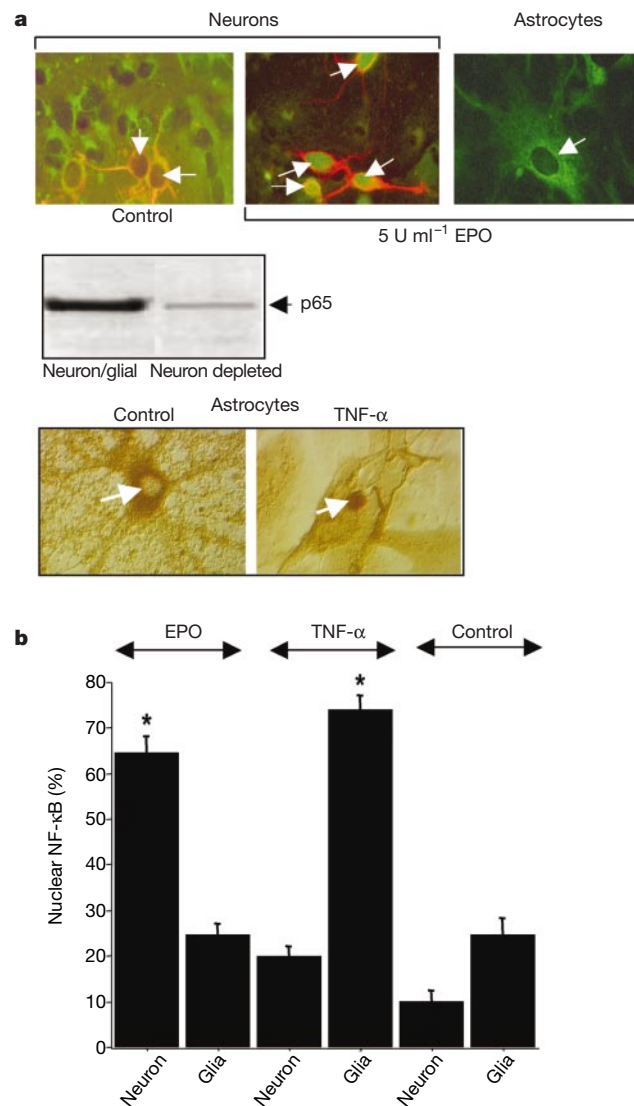


Figure 3 EPO induces nuclear translocation of NF- κ B in neurons rather than astrocytes. **a**, Mixed neuronal/glia cerebrocortical cultures were exposed to EPO (5 U ml⁻¹). Top: anti-NF- κ B (p65 subunit, green) labelling was localized in the neuronal cytoplasm in controls but predominantly in the neuronal nucleus within minutes of EPO exposure. Neurons were identified with anti-MAP-2 antibody (red). Astrocytes did not manifest nuclear accumulation of p65 (arrows point to nuclei), although this subunit was present in the

cytoplasm. Middle: neuron-specific increases in nuclear NF- κ B after EPO were verified on immunoblots of nuclear extracts from mixed or neuron-depleted cultures. Bottom: TNF- α (200 U ml⁻¹) increased nuclear NF- κ B in astrocytes identified with anti-GFAP (glial fibrillary acidic protein, brown). **b**, Quantification of nuclear NF- κ B in neurons versus astrocytes for EPO ($n = 449$ cells), TNF- α ($n = 194$ cells) and control ($n = 352$ cells). Asterisk, $P < 0.001$.

Apoptotic-inhibiting pathways implicated in EPO action involve proteins of the bcl-2 family, extracellular signal-regulated kinases and phosphatidylinositol 3-kinase/Akt^{5,11}. However, additional signalling pathways may be important. For example, although EPO increased bcl-2 family members in our cultures and we have previously reported that these proteins shift the dose-response curve of NO-induced killing¹², they do not offer the high degree of neuroprotection observed here with EPO. Therefore, we sought additional pathways to explain EPO-mediated neuroprotection. Another signalling pathway that is present in many cell types, including neurons and astrocytes, involves the nuclear transcription factor NF- κ B, which is already known to be affected by cytokines, excitotoxins and free radicals, and also to activate bcl-x^{13–16}. Recently, NF- κ B was shown to upregulate transcription of inhibitor-of-apoptosis proteins, including XIAP and c-IAP2, which block activation of specific cell-death proteases (caspases) and subsequent apoptosis¹⁷. NF- κ B protects neurons from excitotoxic and oxidative/nitrosative stress by increasing the activity of Mn- and

Cu,Zn-superoxide dismutases as well as glutathione, thus preventing accumulation of O₂⁻ and ONOO⁻ (refs 18, 19). Hence, we evaluated the potential role of NF- κ B in neurons incubated with EPO. If the neuroprotective effect of EPO were mediated by NF- κ B, EPO should induce nuclear translocation of NF- κ B and its *trans*-activational activity. To test this hypothesis, we initially performed immunoblotting experiments with extracts from cerebrocortical cells incubated with EPO. EPO decreased cytoplasmic and increased nuclear levels of NF- κ B (p65 subunit) in a dose-dependent manner (Fig. 2a, left). Similarly, EPO increased serine-phosphorylated I κ B α in cytoplasmic extracts in a dose-dependent manner (Fig. 2a, right), consistent with NF- κ B activation via phosphorylation-induced degradation of I κ B α .

To determine whether EPO-induced nuclear translocation of NF- κ B is capable of binding DNA, we performed electrophoretic mobility shift assays (EMSA; Fig. 2b, left). Only low levels of activated NF- κ B bound DNA under control conditions. Exposure to neurotoxic concentrations of NO-donor S-nitrosocysteine produced

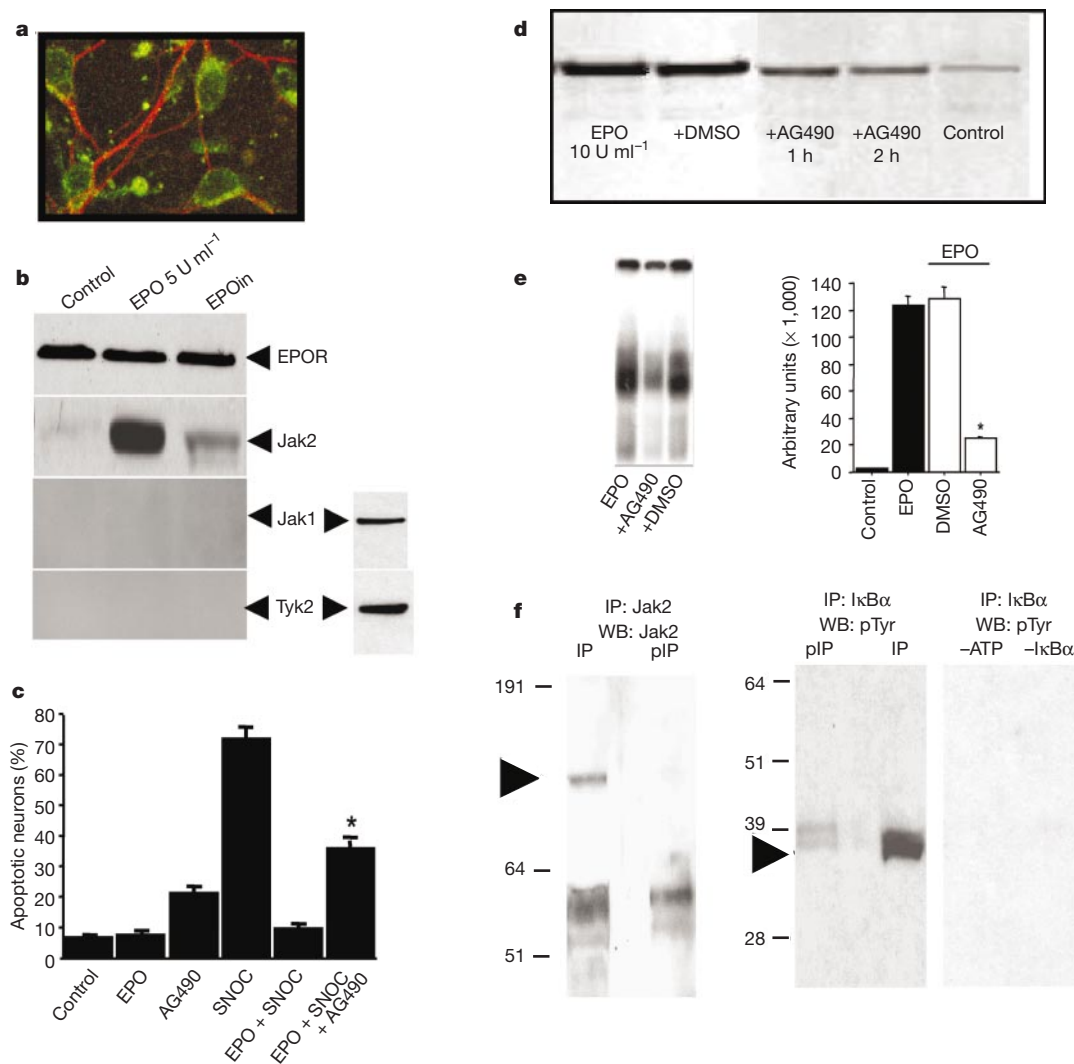


Figure 4 EPO-induced Jak2 and NF- κ B activities in neurons. **a**, In cerebrocortical cultures, Jak2 was detected with a polyclonal antibody (green) and neurons with anti-MAP-2 (red). **b**, Cerebrocortical lysates were immunoprecipitated with anti-EPOR, separated by SDS-PAGE, and immunoblotted with anti-EPOR and anti-Jak2. EPOin, heat-inactivated EPO. **c**, Blocking Jak2 with AG490 (10 μ M) decreased EPO neuroprotection from SNOC. Asterisk, $P < 0.001$. **d**, AG490 reduced EPO-induced nuclear accumulation of NF- κ B (immunoblotted with anti-p65; dimethylsulphoxide (DMSO) diluent for AG490). **e**, EMSA (left) and κ B-dependent reporter gene assay (right) verified that AG490 reduced DNA-binding activity of NF- κ B and NF- κ B *trans*-activational

activity following EPO treatment. **f**, Assay of Jak2 *in vitro* kinase activity. Jak2 was immunoprecipitated (IP), and a sample run on a 10% Bis-Tris gel to insure the presence of Jak2 via western blot (WB: arrow, left panel), as well as the depletion of Jak2 from the post-IP lysate (pIP). Other bands represent immunoglobulin. After incubation of IP Jak2 with exogenous ATP and recombinant I κ B α in an *in vitro* kinase reaction, I κ B α was immunoprecipitated and probed with anti-phosphotyrosine (pTyr) antibody (arrow in middle panel, right lane). As controls, pIP depleted of Jak2 revealed only a faint band (middle panel, left lane), and exogenous ATP or I κ B α was omitted from the reaction (right panel).

small increases in binding of NF- κ B to DNA, which peaked at 1–2 h and was declining by 3 h. In contrast, incubation with increasing concentrations of EPO induced large, sustained (≥ 3 h) increases in binding of NF- κ B to DNA. Thus, EPO resulted in nuclear translocation and DNA binding of NF- κ B. We used a κ B-dependent reporter gene assay to confirm the EMSA results. NF- κ B-dependent transcription was low in control cells, increased after incubation in EPO or TNF- α , which is known to induce κ B-dependent transcription, and was not blocked by S-nitrosocysteine (Fig. 2b, right). Taken together, these results indicate that activation, nuclear translocation and DNA binding of NF- κ B may be involved in EPO-mediated signalling in neurons.

To investigate whether activation of NF- κ B is required for EPO-mediated neuroprotection, in preliminary experiments we used pyrrolidine dithiocarbamate (PDTC), which inhibits NF- κ B activation by preventing its dissociation from I κ B²⁰. In cerebrocortical neurons, S-nitrosocysteine (200 μ M) induced predominantly apoptosis and relatively little necrosis¹⁰. Pre-incubation with EPO significantly reduced the number of apoptotic neurons, but PDTC prevented this neuroprotective effect (Fig. 2c, left). Incubation of S-nitrosocysteine with PDTC did not attenuate apoptosis, indicating that an alternative action of PDTC, to scavenge NO, was not sufficient to prevent neuronal death in this system. In reporter

gene assays, induction of κ B-dependent luciferase activity was greatly reduced by PDTC (Fig. 2c, right). These results indicate that activation of NF- κ B may be required for EPO-mediated neuroprotection, although other actions of PDTC could not be ruled out.

Activated NF- κ B translocates into the nucleus for DNA binding. Therefore, we investigated EPO-induced nuclear translocation of NF- κ B in neurons. Nuclear translocation of NF- κ B is inhibited by the cell-permeable peptide SN50 (ref. 21). Pre-incubation with SN50 (50 μ g ml⁻¹ for 3 h) prevented EPO protection from NO-induced neuronal apoptosis (Fig. 2d, left) and reduced κ B-dependent luciferase activity (Fig. 2d, right). This finding indicates that nuclear translocation of NF- κ B was necessary for the neuroprotective effect of EPO. However, such pharmacological experiments, although suggesting the potential involvement of NF- κ B, have several potential pitfalls; chief among them is the question of specificity of drug action. Therefore, to investigate further the activation of NF- κ B by EPO, we took a molecular approach. Activation of NF- κ B requires its dissociation from I κ B, and this step can be blocked with an I κ B super-repressor that cannot be serine phosphorylated²². We infected cerebrocortical neurons with a recombinant adenovirus encoding an I κ B super-repressor (*Ad5I κ B*) or a control *LacZ* gene (*Ad5LacZ*) that directs the expression of

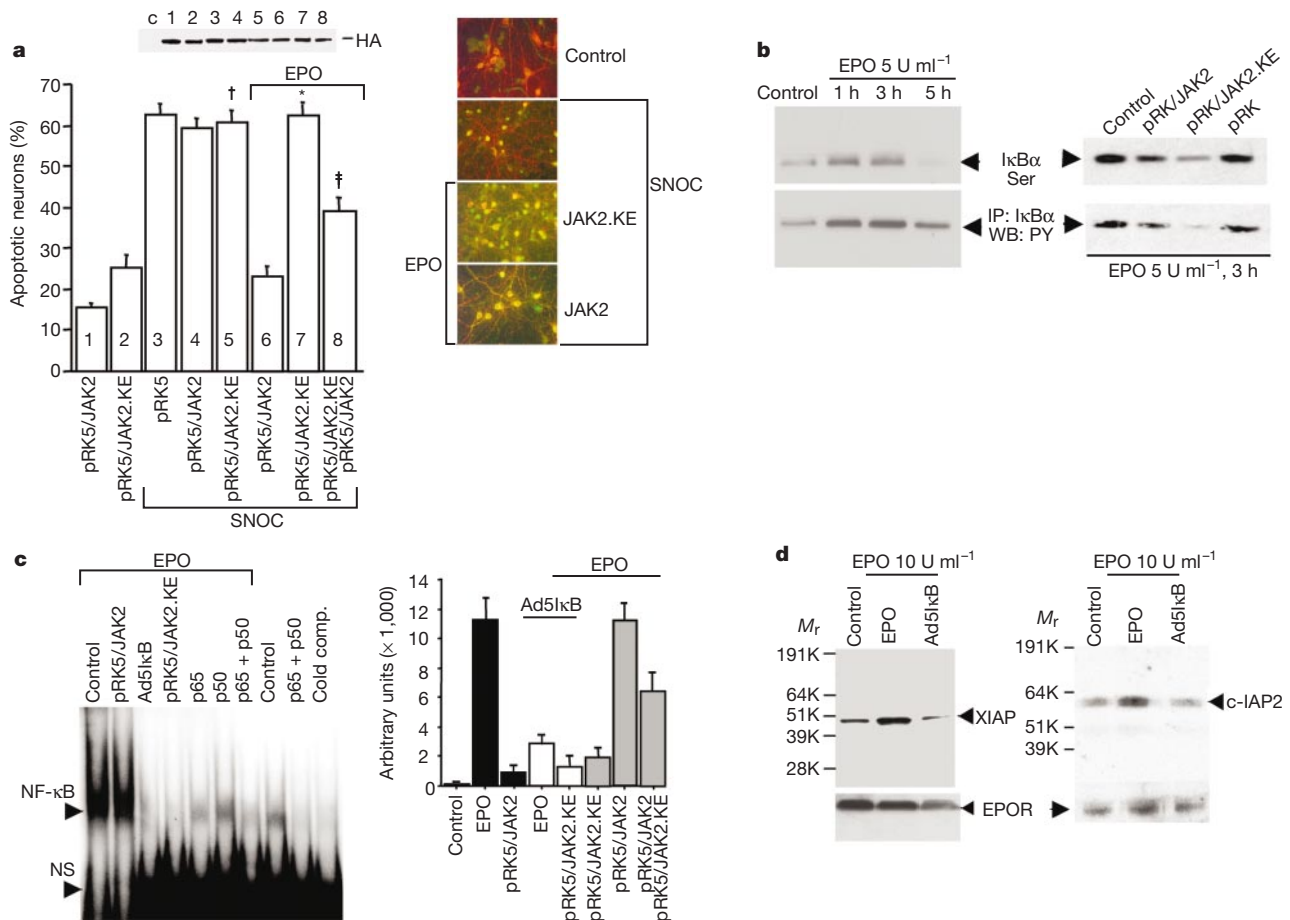


Figure 5 Inhibition of Jak2 function abrogates EPO-mediated neuroprotection. **a**, Cerebrocortical cells were transfected with pRK5/JAK2.KE or wild-type pRK5/JAK2. Western blot of HA-tagged constructs confirmed transfection efficiency (inset). Apoptotic neurons were scored (right, neurons identified with anti-MAP-2 (red) and apoptosis by TUNEL (green)); >400 neurons were counted in a masked fashion for each condition. Cotransfection of JAK2 and JAK2.KE (2:1 ratio) partially restored EPO neuroprotection (dagger symbol, $P < 0.001$ v. pRK5/JAK2.KE; asterisk, $P < 0.001$ v. pRK5/JAK2 + SNOC + EPO; double dagger symbol, $P < 0.001$ v. pRK5/JAK2.KE + SNOC + EPO; all by ANOVA). **b**, EPO-induced phosphorylation of I κ B α detected in whole-cell lysates. After

EPO incubation, cerebrocortical cell lysates were prepared in a kinase buffer. Serine-phosphorylated I κ B α was detected with specific anti-phospho-Ser^{32,36} I κ B α antibody. Tyrosine phosphorylation was detected with a phosphotyrosine antibody (PY) after immunoprecipitation (IP) with anti-I κ B α (left). JAK2.KE, but not JAK2, decreased both serine and tyrosine phosphorylation of I κ B α (right). **c**, Transduction with JAK2.KE or Ad5I κ B inhibited EPO-induced DNA-binding activity of NF- κ B on EMSA (left) and κ B-dependent transactivation activity in reporter gene assays (right). NS, nonspecific binding; Cold comp., competition with cold probe. **d**, EPO-dependent upregulation of XIAP (left) and c-IAP2 (right). M_r , relative molecular mass.

β -galactosidase (β -Gal). Neither Ad5IkB nor Ad5LacZ increased neuronal apoptosis by themselves (Fig. 2e, left). Expression of the IkB super-repressor results in decreased levels of endogenous IkB α because NF- κ B activity is inhibited²². Hence, serine phosphorylation of IkB α induced by both EPO and TNF- α decreased because there was less substrate to phosphorylate (Fig. 2e, right). Additional reporter gene assays showed that expression of Ad5IkB functionally inhibited TNF- α - and EPO-induced κ B-dependent transcriptional activity (Fig. 2f). Moreover, inhibition of NF- κ B activation by Ad5IkB significantly attenuated the protective effect of EPO on NMDA- and NO-induced neuronal apoptosis (Fig. 2g); Ad5LacZ, monitored by anti- β -Gal staining, had no effect. This result is consistent with the hypothesis that NF- κ B activation is necessary for EPO-mediated neuroprotection.

TNF- α also reportedly increases NF- κ B to prevent cell death^{13–15,17–19}. Unlike EPO, however, TNF- α did not protect cerebrocortical neurons from NO-induced apoptosis (Fig. 1c). We hypothesized that these discordant effects were dependent on the cell type in which NF- κ B was activated by EPO or TNF- α in our preparation. To test this premise, we performed NF- κ B immunocytochemistry (Fig. 3a). Under control conditions, NF- κ B was predominantly localized in the cytoplasm of neurons and astrocytes. Within minutes of exposure EPO induced nuclear translocation of NF- κ B primarily in neurons and not in astrocytes. In contrast, TNF- α induced nuclear translocation of NF- κ B predominantly in astrocytes (Fig. 3a, b). Paradoxically, activation of NF- κ B has been reported to mediate neuronal apoptosis and to prevent it^{13–19}. Recently, these seemingly contradictory findings were explained by the fact that acute increases in NF- κ B contribute to an apoptotic signalling pathway, whereas preconditioning stimuli that lead to large increases in steady-state NF- κ B activity provide neuroprotection²³. Accordingly, our preconditioning experiments with EPO led to large increases in NF- κ B activity in neurons, which prevented apoptosis.

We next examined the signalling pathway involved in EPO activation of neuronal NF- κ B. In non-neuronal cells, binding of EPO triggers dimerization of EPORs and activation of Jak2, which phosphorylates tyrosine residues on various substrates, most prominently Stat5 (ref. 24). By immunocytochemistry, we detected Jak2-like protein in rat cerebrocortical neurons and in some non-neuronal cells (Fig. 4a). When we performed immunoprecipitation with anti-EPOR monoclonal antibodies, Jak2 was coprecipitated. Incubation with EPO increased this protein–protein association. In contrast, heat-inactivated EPO as well as several control proteins including bovine serum albumin had no effect. As negative controls, the antibody to EPOR did not immunoprecipitate other tyrosine kinases such as Jak1 and Tyk2 (Fig. 4b, left), although both of these proteins were detectable in the post-immunoprecipitation lysates of cerebrocortical cells (Fig. 4b, right).

To investigate the possible importance of EPO-activated Jak2 to NF- κ B signalling, we examined the effect of inhibiting phosphorylation activity of Jak2 on NF- κ B levels in neuronal nuclei. In preliminary experiments, we used the tyrphostin AG490, a potent inhibitor of Jak2-catalysed phosphorylation²⁵. EPO-mediated neuroprotection from NO-induced apoptosis was significantly blocked when cells were incubated in AG490 (Fig. 4c). We further studied this effect in nuclear extracts from cerebrocortical neurons incubated in EPO. AG490 decreased nuclear translocation of NF- κ B (Fig. 4d). EMSA confirmed these results, demonstrating decreased NF- κ B levels in nuclei after incubation with AG490 (Fig. 4e, left). In reporter gene assays, AG490 blocked EPO-induced κ B-dependent luciferase activity (Fig. 4e, right). Next, we used an *in vitro* kinase assay to test whether Jak2 could directly phosphorylate IkB α , leading to its dissociation from and activation of NF- κ B. In this assay, immunoprecipitated Jak2 from cerebrocortical cells previously treated with EPO induced tyrosine phosphorylation of full-length exogenous, recombinant IkB α within minutes (Fig. 4f).

Nonetheless, serine phosphorylation of IkB α may predominate in NF- κ B activation because inhibition of serine phosphorylation with the IkB super-repressor largely prevented the neuroprotective effect of EPO (see Fig. 2a, g and below).

Pharmacological agents that inhibit Jak2 activity, such as AG490, are limited by questions of specificity. To target Jak2 more specifically, we transfected cerebrocortical neurons with a plasmid containing a gene encoding a kinase-negative mutant Jak2 (pRK5/JAK2.KE) or wild-type Jak2 (pRK5/JAK2)²⁶. JAK2.KE is a dominant interfering form of Jak2; it contains a point mutation that inhibits kinase (phosphorylation) activity, although it can still bind to the EPOR, thus outcompeting endogenous Jak2 (refs 26, 27). Cells were cotransfected with a plasmid that directs expression of enhanced green fluorescent protein (EGFP) so that transfected neurons could be easily identified. Transfection with JAK2.KE, but not JAK2, abrogated the protective effect of EPO against NO-induced neuronal apoptosis (Fig. 5a). Overexpression of wild-type JAK2 was able to overcome the effect of JAK2.KE, indicating that its apoptotic action was not due to nonspecific toxicity. Using specific antibodies, we observed that EPO induced both serine and tyrosine phosphorylation of IkB α ; serine phosphorylation decreased after 5 h whereas levels of tyrosine-phosphorylated IkB α remained relatively constant (Fig. 5b, left). Transfection with JAK2.KE, but not JAK2 or control plasmid, decreased both EPO-induced serine and tyrosine phosphorylation of IkB α (Fig. 5b, right). Additionally, expression of dominant negative Jak2 (JAK2.KE) or IkB super-repressor (Ad5IkB) inhibited EPO-induced NF- κ B activation (Fig. 5c, left). Supershift analysis of DNA–NF- κ B complexes revealed their composition of p65/p50 of NF- κ B subunits. Diminished transcriptional activity in reporter gene assays from cells expressing JAK2.KE or Ad5IkB confirmed the results from EMSA (Fig. 5c, right). Importantly, cotransfection of pRK5/JAK2 (ratio 2:1) partially reversed this action of pRK5/JAK2.KE, indicating the specificity of the effect. Finally, pre-incubation in EPO increased the inhibitor-of-apoptosis gene products XIAP and c-IAP2 on immunoblots of cerebrocortical cells. Ad5IkB blocked EPO-mediated upregulation of XIAP or c-IAP2 but did not change basal levels of expression (Fig. 5d). Taken together, these results firmly indicate that EPO can regulate NF- κ B activity leading to neuroprotection, and that this pathway is mediated by signalling via Jak2.

In summary, we show that preconditioning with EPO protects neurons from excitotoxin- and NO-induced apoptosis, and that NF- κ B participates in this EPO-induced neuroprotection (see Supplementary Information). A short exposure to hypoxia–ischaemia with subsequent reperfusion (termed hypoxic–ischaemic preconditioning) is known to be neuroprotective from a subsequent, more prolonged hypoxic–ischaemic insult²⁸. EPO is induced by a hypoxic stimulus via HIF-1. Hence, EPO may contribute to the neuroprotective effect of hypoxic–ischaemic preconditioning and may have potential therapeutic value in the brain. □

Methods

Preparation of EPO

Recombinant human EPO was of the highest purity (>99.999%; AMGEN).

Apoptosis analysis, plasmid transduction and NO

For apoptosis detection, 18 h after insult TUNEL (Apoptosis Detection Kit with dUTP labelled with fluorescein/green (Promega) or diethylaminocoumarin/blue (New England Nuclear)) was used in conjunction with condensation of nuclear morphology. Plasmids were introduced 24 h before experimentation via adenoviral construct²² or lipid-based transfection (LipofectAMINE 2000, Gibco). As an index of NO levels, nitrite concentration was monitored with the Griess reaction²⁹.

Immunocytochemistry and *in vitro* kinase assays

Cells in control or neuron-depleted cultures were lysed in ice-cold buffer (50 mM Tris-Cl buffer at pH 8.0, 150 mM NaCl, 100 μ g ml⁻¹ phenylmethyl sulphonyl fluoride, 1 μ g ml⁻¹ aprotinin, 1% Triton X-100), separated by SDS–PAGE and immunoblotted with anti-EPOR, anti-Jak2, anti-phosphoserine or anti-phosphotyrosine antibodies²¹. In some cases, lysates were first immunoprecipitated with anti-EPOR before western blotting.

To evaluate tyrosine phosphorylation of I κ B α ³⁰ after treatment of cells with EPO, lysates were immunoprecipitated with anti-I κ B α and then probed with anti-phosphotyrosine antibodies. To detect serine phosphorylation of I κ B α , lysates were immunoblotted without immunoprecipitation and probed directly with antibodies specific for phosphorylation of I κ B α serine residues 32 and 36 (Santa Cruz Biotechnology). For Jak2 *in vitro* kinase assays, after cell lysis Jak2 was immunoprecipitated with 20 μ g of a polyclonal antibody (Santa Cruz Biotechnology). Jak2 was then resuspended in kinase buffer (including 25 mM HEPES, 25 mM MgCl₂, 0.1 mM Na-orthovanadate and 2 mM dithiothreitol) plus 10 μ M ATP and 30 μ g full-length recombinant I κ B α . After reaction for 30 min, anti-I κ B α immune complexes were resolved by SDS–PAGE, probed with anti-phosphotyrosine antibodies, and visualized by ECL (Amersham).

Electrophoretic mobility-shift assays (EMSA)

Nuclear extracts were obtained from cerebrocortical cultures¹⁹. Binding of NF- κ B to DNA was assayed with a double-stranded probe labelled with ³²P-dUTP that binds to the consensus sequence (Santa Cruz Biotechnology). Nuclear lysates were incubated with the labelled probe for 2 h at 37 °C, resolved on a 7% native polyacrylamide gel, and exposed to X-ray film²¹. Mutated probe was used as a control. Antibodies specific for p50 and p65 NF- κ B subunits were used for supershift analysis. In non-neuronal cells, S-nitrosylation (transfer of NO-related species to a critical thiol from S-nitrosocysteine or other donors) has been reported to block DNA binding by NF- κ B. However, this phenomenon did not affect the EMSA results reported here in neurons because S-nitrosocysteine did not prevent EPO-induced binding.

Reporter gene assays

Cerebrocortical cells were transfected with pNF κ B-Luc using calcium phosphate precipitation (Stratagene). Two days later, cells were lysed and mixed with luciferase assay reagent (Promega), and the activity was measured in a luminometer. All measurements were normalized against a non- κ B-dependent control plasmid, pCIS-CK (Stratagene). Results represent mean of three experiments measured in triplicate.

Received 23 January; accepted 8 May 2001.

1. Digicaylioglu, M. *et al.* Localization of specific erythropoietin binding sites in defined areas of the mouse brain. *Proc. Natl Acad. Sci. USA* **92**, 3717–3720 (1995).
2. Masuda, S. *et al.* Functional erythropoietin receptors of the cells with neuronal characteristics—comparison with receptor properties from erythroid cells. *J. Biol. Chem.* **268**, 11208–11216 (1993).
3. Bernaudin, M. *et al.* Neurons and astrocytes express EPO mRNA: oxygen-sensing mechanisms that involve the redox-state of the brain. *Glia* **30**, 271–278 (2000).
4. Morishita, E., Masuda, S., Nagao, M. & Sasaki, R. Erythropoietin receptor is expressed in rat hippocampal cerebral cortical neurons, and erythropoietin prevents *in vitro* glutamate-induced neuronal death. *Neuroscience* **76**, 105–116 (1997).
5. Sirén, A.-L. *et al.* Erythropoietin prevents neuronal apoptosis after cerebral ischemia and metabolic stress. *Proc. Natl Acad. Sci. USA* **98**, 4044–4049 (2001).
6. Anagnostou, A., Lee, E. S., Kessimian, N., Levinson, R. & Steiner, M. Erythropoietin has a mitogenic and positive chemotactic effect on endothelial cells. *Proc. Natl Acad. Sci. USA* **87**, 5978–5982 (1990).
7. Parganas, E. *et al.* Jak2 is essential for signaling through a variety of cytokine receptors. *Cell* **93**, 385–395 (1998).
8. Lipton, S. A. *et al.* A redox-based mechanism for the neuroprotective and neurodestructive effects of nitric oxide and related nitroso-compounds. *Nature* **364**, 626–632 (1993).
9. Dawson, V. L., Dawson, T. M., Bartley, D. A., Uhl, G. R. & Snyder, S. H. Mechanisms of nitric oxide-mediated neurotoxicity in primary brain cultures. *J. Neurosci.* **13**, 2651–2661 (1993).
10. Bonfoco, E., Krainc, D., Ankarcrona, M., Nicotera, P. & Lipton, S. A. Apoptosis and necrosis: two distinct events induced, respectively, by mild and intense insults with N-methyl-D-aspartate or nitric oxide/superoxide in cortical cell cultures. *Proc. Natl Acad. Sci. USA* **92**, 7162–7166 (1995).
11. Gregory, T. *et al.* GATA-1 and erythropoietin cooperate to promote erythroid cell survival by regulating bcl-x_i expression. *Blood* **94**, 87–96 (1999).
12. Bonfoco, E. *et al.* Bcl-2 delays apoptosis and PARP cleavage induced by NO donors in GT1-7 cells. *Neuroreport* **8**, 273–276 (1996).
13. Beg, A. A. & Baltimore, D. An essential role for NF- κ B in preventing TNF- α -induced cell death. *Science* **274**, 782–784 (1996).
14. Wang, C. Y., Mayo, M. W. & Baldwin, A. S. Jr TNF- and cancer therapy-induced apoptosis: potentiation by inhibition of NF- κ B. *Science* **274**, 784–787 (1996).
15. Van Antwerp, D. J., Martin, S. J., Kafri, T., Green, D. R. & Verma, I. M. Suppression of TNF- α -induced apoptosis by NF- κ B. *Science* **274**, 787–789 (1996).
16. Grilli, M., Pizzi, M., Memo, M. & Spano, P. Neuroprotection by aspirin and sodium salicylate through blockade of NF- κ B activation. *Science* **274**, 1383–1385 (1996).
17. Wang, C. Y., Mayo, M. W., Korneluk, R. G., Goeddel, D. V. & Baldwin, A. S. Jr NF- κ B antiapoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation. *Science* **281**, 1680–1683 (1998).
18. O'Neill, L. A. & Kaltschmidt, C. NF- κ B: a crucial transcription factor for glial and neuronal cell function. *Trends Neurosci.* **20**, 252–258 (1997).
19. Mattson, M. P., Goodman, Y., Luo, H., Fu, W. & Furukawa, K. Activation of NF- κ B protects hippocampal neurons against oxidative stress-induced apoptosis: evidence for induction of manganese superoxide dismutase and suppression of peroxynitrite production and protein tyrosine nitration. *J. Neurosci.* **19**, 681–697 (1999).
20. Schreck, R., Meier, B., Mannel, D. N., Droge, W. & Baeuerle, P. A. Dithiocarbamates as potent inhibitors of nuclear factor κ B activation in intact cells. *J. Exp. Med.* **175**, 1181–1194 (1992).
21. Lin, Y. Z., Yao, S. Y., Veach, R. A., Torgerson, T. R. & Hawiger, J. Inhibition of nuclear translocation of transcription factor NF- κ B by a synthetic peptide containing a cell membrane-permeable motif and nuclear localization sequence. *J. Biol. Chem.* **270**, 14255–14258 (1995).
22. Iimuro, Y. *et al.* NF- κ B prevents apoptosis and liver dysfunction during liver regeneration. *J. Clin. Invest.* **101**, 802–811 (1998).

23. Lezoualc'h, F., Sagara, Y., Holsboer, F. & Behl, C. High constitutive NF- κ B activity mediates resistance to oxidative stress in neuronal cells. *J. Neurosci.* **18**, 3224–3232 (1998).
24. Ihle, J. N., Witthuhn, B. A., Quelle, F. W., Yamamoto, K. & Silvennoinen, O. Signaling through the hematopoietic cytokine receptors. *Annu. Rev. Immunol.* **13**, 369–398 (1995).
25. Meydan, N. *et al.* Inhibition of acute lymphoblastic leukaemia by a Jak-2 inhibitor. *Nature* **379**, 645–648 (1996).
26. Briscoe, J. *et al.* Kinase-negative mutants of JAK1 can sustain interferon-gamma-inducible gene expression but not an antiviral state. *EMBO J.* **15**, 799–809 (1996).
27. Zhuang, H. *et al.* Inhibition of erythropoietin-induced mitogenesis by a kinase-deficient form of Jak2. *J. Biol. Chem.* **269**, 21411–21414 (1994).
28. Gage, A. T. & Stanton, P. K. Hypoxia triggers neuroprotective alterations in hippocampal gene expression via a heme-containing sensor. *Brain Res.* **719**, 172–178 (1996).
29. Schmidt, H. H. W. & Kelm, M. in *Methods in Nitric Oxide Research* (eds Feelisch, M. & Stamler, J. S.) 491–497 (Wiley, Chichester, 1996).
30. Imbert, V. *et al.* Tyrosine phosphorylation of I κ B- α activates NF- κ B without proteolytic degradation of I κ B- α . *Cell* **86**, 787–798 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Kaul, N. Moayeri, B. Price, M. Cokol and M. Altinoz for insightful discussions or technical advice, and the Genetics Institute, Cambridge, Massachusetts, for supplying the anti-EPOR monoclonal antibodies. The complementary DNA strands for the I κ B super-repressor (Ad5I κ B) and kinase-negative mutant Jak2 (JAK2.KE) were the gifts of R. R. Ratan and J. Ihle, respectively. This work was supported in part by grants from the National Institutes of Health and American Heart Association (S.A.L.).

Correspondence and requests for materials should be addressed to S.A.L. (email: slipton@burnham.org).

..... Spred is a Sprouty-related suppressor of Ras signalling

Toru Wakioka^{*†‡}, Atsuo Sasaki^{*}, Reiko Kato^{*†‡}, Takanori Shouda^{*}, Akira Matsumoto^{*}, Kanta Miyoshi^{*†‡}, Makoto Tsuneoka[‡], Setsuro Komiya[†], Roland Baron[§] & Akihiko Yoshimura^{*†‡}

^{*} Division of Molecular and Cellular Immunology, Medical Institute of Bioregulation, Kyushu University, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

[†] Department of Orthopaedic Surgery, Faculty of Medicine, Kagoshima University, Kagoshima 899, Japan

[‡] Institute of Life Science, Kurume University, Kurume 839-0861, Japan

[§] Department of Cell Biology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

Cellular proliferation, and differentiation of cells in response to extracellular signals, are controlled by the signal transduction pathway of Ras, Raf and MAP (mitogen-activated protein) kinase. The mechanisms that regulate this pathway are not well known. Here we describe two structurally similar tyrosine kinase substrates, Spred-1 and Spred-2. These two proteins contain a cysteine-rich domain related to Sprouty (the SPR domain) at the carboxy terminus. In *Drosophila*, Sprouty inhibits the signalling by receptors of fibroblast growth factor (FGF) and epidermal growth factor (EGF) by suppressing the MAP kinase pathway^{2–7}. Like Sprouty, Spred inhibited growth-factor-mediated activation of MAP kinase. The Ras–MAP kinase pathway is essential in the differentiation of neuronal cells and myocytes. Expression of a dominant negative form of Spred and Spred-antibody microinjection revealed that endogenous Spred regulates differentiation in these types of cells. Spred constitutively associated with Ras but did not prevent activation of Ras or membrane translocation of Raf. Instead, Spred inhibited the activation of MAP kinase by suppressing phosphorylation and activation of Raf. Spred may represent a

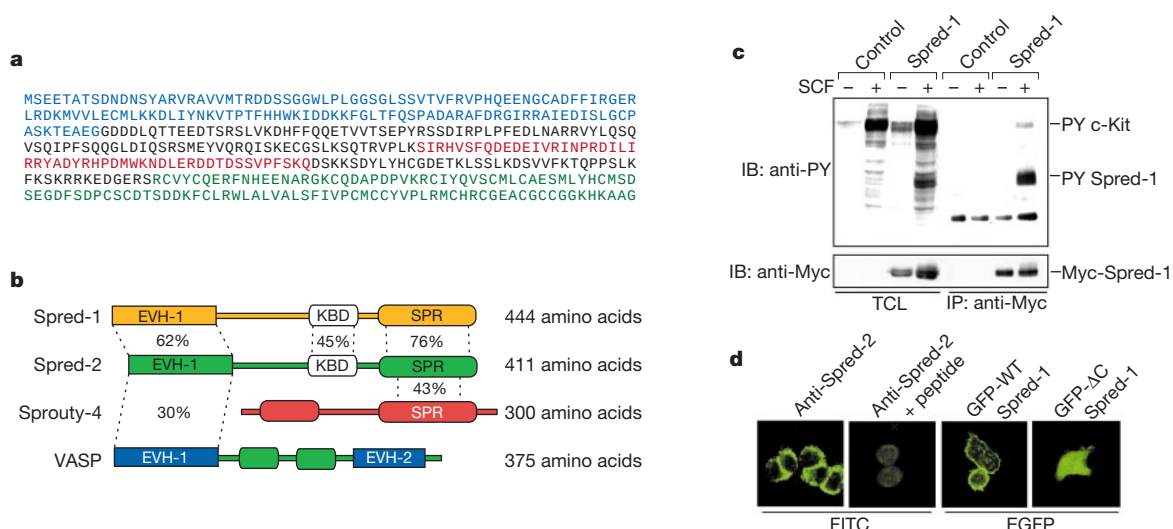


Figure 1 Characterization of Spred molecules. **a**, Amino acid sequence of Spred-1. The EVH-1 domain is shown in blue, the KBD in red, and the SPR domain in green. **b**, Comparison of the domain structure of murine Spred-1 and -2 with that of murine Sprouty-4 and VASP. **c**, Tyrosine phosphorylation (PY) of Myc-tagged Spred-1 expressed in 293 cells in response to SCF. IB, immunoblot; IP, immunoprecipitate; TCL, total cell

lysates. **d**, Subcellular localization of Spred. The endogenous Spred-2 protein in PC12 cells was detected using immunofluorescence microscopy without or with antigen peptide. FITC, fluorescein isothiocyanate. The two right panels show a fluorescent microscopic view of wild-type (GFP-WT) and C-terminal truncated (GFP-ΔC) Spred-1 fused to EGFP in PC12 cells.

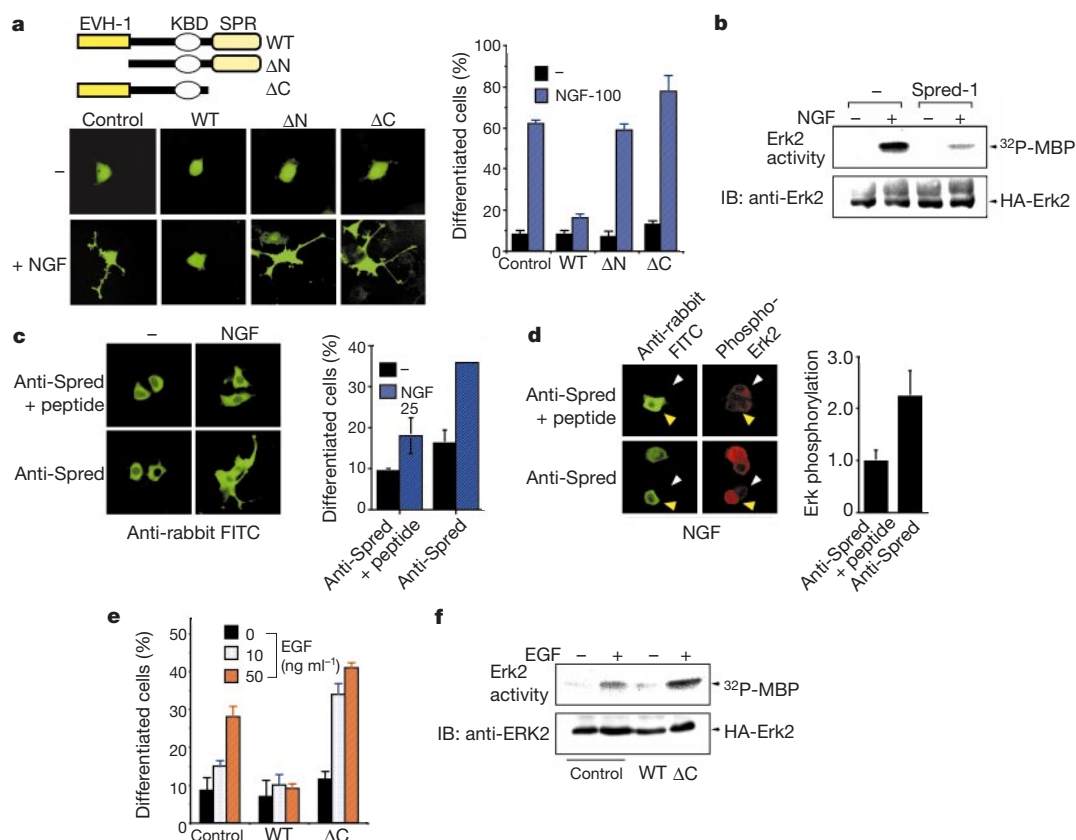


Figure 2 Effects of Spred on differentiation and Erk2 activity of PC12 cells. **a**, Wild-type (WT) and N-terminal- (ΔN) or C-terminal-truncated (ΔC) Spred-1 constructs were expressed in PC12 cells with EGFP, then incubated with 100 ng ml⁻¹ NGF for 3 d. **b**, PC12 cells transfected with expression vectors encoding wild-type Spred-1 and HA-tagged Erk2 were stimulated with NGF for 30 min. HA-Erk2 was immunoprecipitated and an *in vitro* kinase assay was performed with MBP as the substrate. IB, immunoblot. **c**, **d**, PC12 cells were microinjected with a mixture of affinity-purified anti-Spred-1 and -2 antibodies, with

or without antigen peptide, and then treated with 25 ng ml⁻¹ NGF for 2 d (**c**) or 5 min (**d**). Activated Erk was visualized with a monoclonal anti-phosphorylated-Erk antibody and quantified. The y axis shows relative fluorescence intensity of each cell. In **d**, yellow arrowheads indicate injected cells, and white arrowheads uninjected cells. **e**, **f**, Effects of dominant negative Spred. Cells were co-transfected with ΔC-Spred-1 mutant and EGF receptor, then treated with 10 or 50 ng ml⁻¹ EGF for 3 d (**e**) or 30 min (**f**).

class of proteins that modulate Ras–Raf interaction and MAP kinase signalling.

From an osteoclast complementary DNA library, we isolated a tyrosine-kinase-binding protein by a yeast two-hybrid system using the c-Kit and c-Fms tyrosine kinase domains as bait. Full-length cDNA of this gene encodes a protein that contains a C-terminal SPR domain and an amino-terminal Ena/Vasodilator-stimulated phosphoprotein (VASP) homology-1 (EVH-1) domain¹ (Fig. 1a). Thus, we named this gene *Spred-1*: Sprouty-related protein with EVH-1 domain. We found a very similar gene (*Spred-2*) in the database and cloned its full-length cDNA (Fig. 1b). Using various deletion mutants, we identified a c-Kit-binding domain (KBD) composed of about 50 amino acids (codon 234–286) (see Supplementary Information Fig. 1). This region is not related to any previously identified tyrosine kinase interaction domains such as SH2, PTB or c-Met-binding domain. *Spred-1* was tyrosine phosphorylated in response to stem cell factor (SCF), platelet-derived growth factor (PDGF) and EGF (Fig. 1c and data not shown), and efficient phosphorylation of *Spred-1* required the KBD region (Supplementary Information Fig. 2). Using immunofluorescence microscopy, we detected localization of endogenous *Spred-2* to the plasma membrane (Fig. 1d). Membrane localization of *Spred* was confirmed by exogenously expressed *Spred* fused to enhanced green fluorescent protein (EGFP) (Fig. 1d). The C-terminal SPR domain was essential for plasma membrane localization, as a deletion

mutant lacking SPR domain (GFP- Δ C) localized in the cytoplasm (Fig. 1d).

Because *Spred* contains a Sprouty-related domain, we examined the effect of *Spred* on the Ras–MAP kinase pathway. First, we examined the differentiation of PC12 pheochromocytoma cells, which is induced by nerve growth factor (NGF) and dependent on MAP kinase⁸. As shown in Fig. 2a, overexpression of *Spred-1* or *Spred-2* strongly inhibited NGF-induced differentiation of PC12 cells. Both EVH-1 and SPR domains were essential for the suppression of differentiation (Fig. 2a, see Δ N and Δ C). *Spred-1* suppressed NGF-induced activation of Erk2 MAP kinase (Fig. 2b). Next, we inhibited endogenous *Spred* by microinjecting affinity-purified anti-*Spred-1* and -*Spred-2* antibodies. The microinjection augmented neurite outgrowth of PC12 cells treated with a low concentration of NGF (Fig. 2c) as well as NGF-induced Erk2 activation (Fig. 2d). These results indicate that the endogenous *Spred* proteins suppress growth-factor-induced activation of MAP kinase and reduce the threshold of growth factor sensitivity for differentiation in PC12 cells. We noticed that the C-terminal deletion mutant augmented NGF-induced neurite outgrowth of PC12 cells (Fig. 2a), suggesting that the Δ C mutant may function as a dominant negative form against endogenous *Spred* proteins. Indeed, overexpression of the Δ C mutant augmented differentiation of PC12 cells treated with a low concentration (10 ng ml⁻¹) of EGF (Fig. 2e) and enhanced EGF-induced activation of Erk2 (Fig. 2f).

The negative effects of *Spred* on the Ras–MAP kinase pathway were confirmed in a different system. C2C12 cells differentiated into myotubes when cultured in 2% horse serum (HS) (differentiation medium) for 3–5 days. MAP kinase activity has been shown to drop rapidly after switching to differentiation medium, but recovers after 3 days^{9,10}. This decrease in MAP kinase activity in differentiation medium is essential for myotube formation. The levels of *Spred-1* in C2C12 cells had increased by day 2 and dropped sharply on day 3 (Fig. 3a), and were inversely correlated to the levels of MAP kinase activity¹⁰. As shown in Fig. 3b and c, forced expression of dominant negative Ras (N17-Ras) induced morphological changes in C2C12 cells and reduced MAP kinase activity, even under normal growth conditions (20% FBS). Forced expression of wild-type *Spred-1* and -2 exhibited similar effects (Fig. 3b, c). In contrast, C2C12 cells transfected with the constitutively activated form of Raf (Δ N-Raf) or dominant negative mutants (Δ C) of *Spred-1* and -2 did not differentiate into myotubes even in differentiation medium, and a higher level of MAP kinase activity was maintained (Fig. 3d, e). Thus, Δ C mutants inhibited C2C12 cell differentiation probably by augmenting MAP kinase signalling, supporting the hypothesis that endogenous *Spred* proteins function as suppressors of the MAP kinase pathway in various systems.

Next, we investigated the molecular mechanism by which *Spred* suppresses the Ras–MAP kinase pathway. As one of the nuclear targets of MAP kinase is Elk-1, a transcription factor of the Ets family, EGF-induced activation of MAP kinase can be monitored by measuring the rate of Elk-1-dependent transcription¹¹. In 293 cells, forced expression of *Spred-1* or -2 dose-dependently suppressed EGF-dependent Elk-1 activation (Supplementary Information Fig. 3a). The negative effect of *Spred-1* and -2 was comparable to that of Ras GTPase activating protein (rasGAP) and N17-Ras, and *Spred-1* and -2 were more potent inhibitors than was murine Sprouty-4 or the Raf kinase inhibitor protein (RKIP)¹². Both EVH-1 and SPR domains were necessary for the suppression of Elk-1 activation. Replacement of the EVH-1 domain of *Spred-1* with that of Wiskott–Aldrich syndrome protein (WASP) abolished the inhibitory activity of *Spred-1* (Supplementary Information Fig. 3c), suggesting that the EVH-1 domain of *Spred-1* may interact with a specific target required for suppression of the MAP kinase pathway. In contrast, the KBD region was not essential but required for efficient suppression of the MAP kinase pathway (Supplementary Information Figs 2b and 3c).

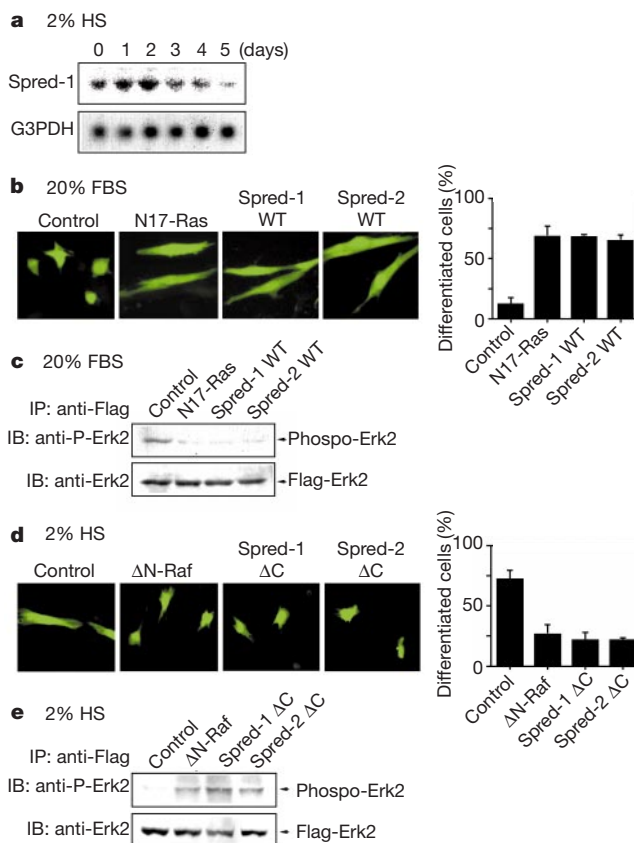


Figure 3 Effects of wild-type (WT) and C-terminal-truncated (Δ C) *Spred* proteins on differentiation of C2C12 cells and Erk2 activation. **a**, C2C12 cells were cultured in differentiation medium (2% HS) for the indicated periods and *Spred-1* mRNA was measured. **b**, **c**, Wild-type *Spred-1*, *Spred-2* and N17-Ras constructs were transfected with EGFP (**b**) or Flag-tagged Erk2 (**c**) then cultured in the growth medium (20% FBS) for 5 d (**b**) or 2 d (**c**). Cell morphology (**b**) and Erk2 phosphorylation (**c**) were measured. **d**, **e**, Δ C *Spred-1* and *Spred-2* constructs and constitutively activated Raf (Δ N-Raf) were transfected with EGFP (**d**) or Flag-tagged Erk2 (**e**). Cells were cultured in differentiation medium (2% HS) for 5 d (**d**) or 2 d (**e**).

We determined which component of the Ras–MAP kinase pathway is suppressed by Spred. Ras directly interacts with and activates Raf. Raf phosphorylates and activates MEK, which in turn phosphorylates and activates MAP kinases. Spred inhibited activation of Elk-1 induced by active Ras (V12-Ras), but not that induced by active MEK or active Raf (Δ N-Raf) (Supplementary Information Fig. 3b). Therefore, the target of Spred is probably located between Ras and Raf. To test this hypothesis, we examined the effect of Spred-1 on EGF-induced Ras and Raf activation (Fig. 4a). Interestingly, Spred sustained Ras activation, whereas it inhibited Raf activation, as measured by autophosphorylation (Fig. 4a) and by *in vitro* kinase assay (Fig. 4b). Furthermore, like rasGAP, Spred inhibited the phosphorylation of Raf on Ser 338, which is required for Raf activation, but not on Ser 259, which is not (refs 13, 14; Fig. 4c). Thus, Spred inhibits MAP kinase activity by suppressing Raf activation. In contrast, Spred did not affect EGF- or V12-Ras-dependent Akt activation, or EGF-dependent tyrosine phosphorylation of phospholipase-C γ and membrane ruffling induced by the small GTP-binding protein Rac¹⁵ (Supplementary Information Fig. 4). These data suggest that Spred specifically suppresses the Ras–Raf signalling pathway.

We then examined the interaction between Raf, Ras and Spred in C2C12 cells. After immunoprecipitation, endogenous Spred-2 was identified as a protein with a relative molecular mass of 40,000 (M_r

40K) that co-immunoprecipitated with endogenous Ras but not Raf (Fig. 5a). Furthermore, Ras and Spred co-localized at the plasma membrane, independent of EGF stimulation (Fig. 5b). Unexpectedly, Raf was translocated into the plasma membrane even in the presence of Spred-1 and co-localized with Spred-1 (Fig. 5b), thus indicating that inhibition of Raf activation by Spred is not due to a simple masking of the effector domain of Ras. We therefore examined the effect of Spred on the interaction between Raf and activated Ras (V12-Ras). As shown in Fig. 5c, the amount of Raf co-precipitated with active Ras was increased by Spred-1 co-expression. Augmentation of the interaction between Raf and Ras by Spred was confirmed by monitoring the time course of EGF-induced translocation of Raf to the plasma membrane. Raf was retained longer in the plasma membrane by overexpression of Spred-1 or Spred-2 (Supplementary Information Fig. 5). Thus Spred proteins potentiate Ras–Raf interaction, and Raf may not be accessible to Raf kinase

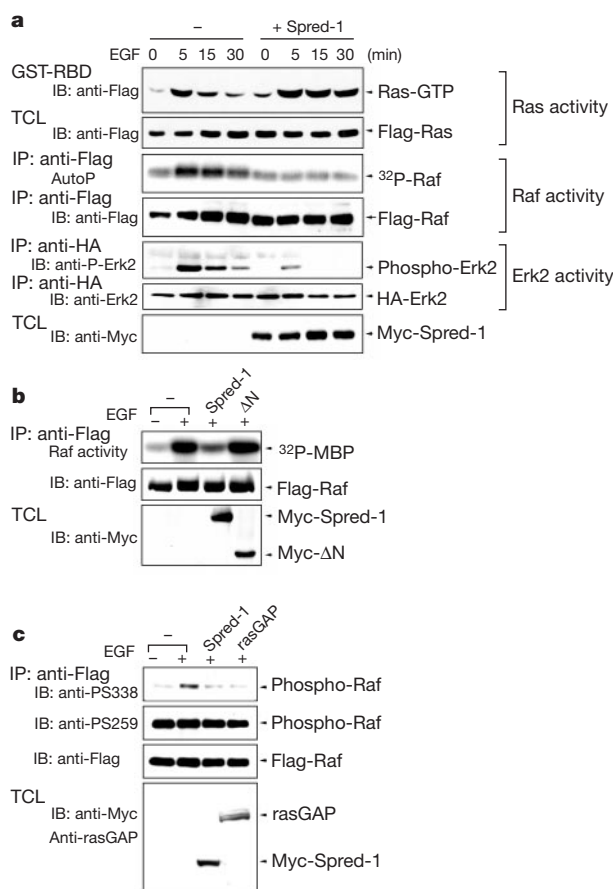


Figure 4 Spred inhibits the activation of MAP kinase by suppressing Raf activation. **a**, Time course of Erk2, Ras and Raf activation in 293 cells transfected with HA-Erk2, wild-type Flag-Ras and Flag-Raf with or without Spred-1. Activated Ras was precipitated with GST-Ras binding domain (GST-RBD) beads, and Raf activation was assessed by an *in vitro* autophosphorylation (AutoP) assay. **b**, Suppression of Raf kinase activity confirmed by an *in vitro* kinase assay using MAP kinase cascade components and MBP. **c**, Effect of Spred-1 on Raf phosphorylation. Flag-Raf was immunoprecipitated (IP) and blotted (IB) with anti-phospho-Ser 338 (PS338) or Ser 259 (PS259) Raf antibodies.

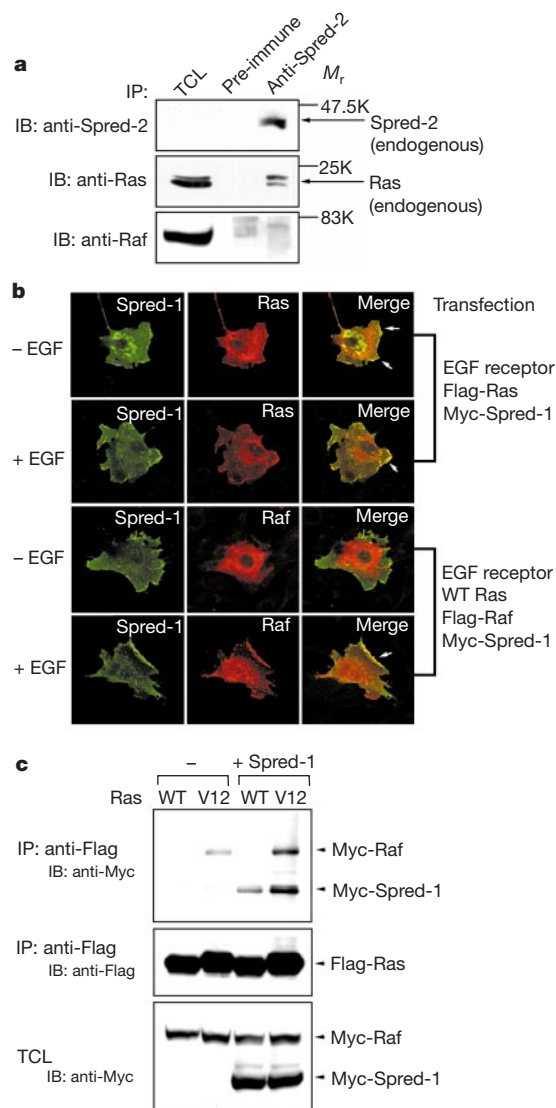


Figure 5 Interaction of Spred with Ras and Raf. **a**, Anti-Spred-2 immunoprecipitates (IP) from C2C12 cells were blotted (IB) with the indicated antibodies. TCL, total cell lysates. **b**, C2C12 cells were transfected with the EGF receptor, Myc-Spred-1 and Flag-Ras, or Myc-Spred-1, Flag-Raf and wild-type (WT) Ras. After stimulation with EGF for 5 min, Myc-tagged and Flag-tagged proteins were stained with secondary antibodies conjugated with FITC (green) and Cy3 (red), respectively. Arrows indicate co-localization. **c**, C2C12 cells were transfected with Flag-tagged wild-type (WT) or constitutively activated (V12) Ras, Myc-tagged Spred and Myc-tagged Raf. Anti-Flag immunoprecipitates were blotted with anti-Myc and anti-Flag antibodies.

when associated with the Spred–Ras complex, thereby preventing its activation.

Drosophila Sprouty is induced by FGF and inhibits Ras–Raf activation by interacting with rasGAP and Drk/Grb2 (ref. 3). The Spred orthologue of *Drosophila* is AE33, which was cloned as a probable target of the *rough* transcription factor that regulates photoreceptor cell development¹⁶. Thus, Spred and Sprouty are evolutionarily conserved, and several isoforms have been identified in mammals. Spred and Sprouty may negatively regulate the Ras–MAP kinase pathway in response to different stimuli in different tissues or organs. □

Methods

Screening and cloning of Spred-1 and -2

Randomly primed cDNA synthesized from mouse osteoclast messenger RNA was cloned into λ phage pASV3A vector and packaged. Plasmids were recovered by infecting phage into *Escherichia coli* BNN-132 (ref. 17). The yeast two-hybrid screen was done as described¹⁸.

Plasmid construction

Mouse Spred-1 and -2, human Ras, Raf and MEK, and murine Sprouty-4 cDNAs were cloned into pcDNA3 with a six-repeated Myc tag or pCMV2 with a Flag tag at the N terminus. Spred-1 and -2 were also cloned into pEGFP (Clontech) to introduce the EGFP tag at the N terminus. Erk2 MAP kinase tagged with haemagglutinin (HA) and Flag was a gift of Y. Goto. WASP cDNA was donated by H. Miki and T. Takenawa. Deletion, substitution and chimeric mutants were generated by standard polymerase chain reaction, as described¹⁹.

Differentiation of PC12 and C2C12 cells

Rat pheochromocytoma-derived PC12 cells were maintained in DMEM supplemented with 10% FBS and 5% HS. We transfected PC12 cells seeded in plates coated with poly-L-lysine with EGFP-tagged Spred-1, EGF receptor and HA-tagged Erk2, using TransFast (Promega). Myc-tagged Spred-1 was introduced into PC12 cells using a retrovirus vector (pMX-IRES-EGFP) carrying Spred-1 cDNA and IRES-EGFP²⁰. From 48 h after transfection or infection, PC12 cells were cultured in the presence of NGF or EGF for 72 h, then examined by fluorescence microscopy. Cells with processes longer than 1.5 times the diameter of the cell body were considered to be positive for neurite outgrowths⁸. Affinity-purified Spred antibodies with or without antigen peptides were microinjected with Eppendorf transjector 5246 or 5171 and examined by inverted microscopy (Zeiss), as described²¹. *In situ* Erk phosphorylation assay was performed as described¹².

Differentiation of C2C12 cells and transfection of Spred-1 were as described⁹. Cells grown in 24-well plates were co-transfected with pEGFP empty vector and Flag-tagged Spred proteins or their deleted mutants, Myc-tagged N17-Ras or Δ N-Raf, using LipofectAMINE (Gibco BRL). After transfection, cells were cultured for 5 d in 20% FBS or 2% HS, then examined by fluorescence microscopy. The elongated and multinucleated myotube-like cells were considered to be differentiated.

Immunochemical analysis

Immunoprecipitation and immunoblotting were performed with anti-Myc (9E10), anti-Flag (M2), anti-Ras (Calbiochem), anti-Erk2 (Santa Cruz), anti-HA (12CA5), anti-phosphorylated-Erk2 (Promega), anti-phospho-Ser 338 (Upstate) or Ser 259 Raf (New England) antibodies, as described²². The Ras-GTP form was precipitated with the Raf N-terminal Ras-binding domain fused to GST, as described²³. Anti-Spred-1 or -2 antibody was prepared by immunizing rabbits with the KLH-conjugated peptides FKSKRRKEDG ERSRC or IKTQPPRAKSKRRKENGEC, respectively.

Kinase assays

In vitro kinase assay for Erk2 and Raf was performed using myelin basic protein as a substrate according to methods described¹².

Received 17 April; accepted 20 June 2001.

- Prehoda, K. E., Lee, D. J. & Lim, W. A. Structure of the enabled/VASP homology 1 domain–peptide complex: a key component in the spatial control of actin assembly. *Cell* **97**, 471–480 (1999).
- Hacohen, N., Kramer, S., Sutherland, D., Hiromi, Y. & Krasnow, M. A. *sprouty* encodes a novel antagonist of FGF signaling that patterns apical branching of the *Drosophila* airways. *Cell* **92**, 253–263 (1998).
- Casci, T., Vinos, J. & Freeman, M. Sprouty, an intracellular inhibitor of Ras signaling. *Cell* **96**, 655–665 (1999).
- de Maximy, A. A. *et al.* Cloning and expression pattern of a mouse homologue of *Drosophila sprouty* in the mouse embryo. *Mech. Dev.* **81**, 213–216 (1999).
- Tefft, J. D. *et al.* Conserved function of *mSpry-2*, a murine homolog of *Drosophila sprouty*, which negatively modulates respiratory organogenesis. *Curr. Biol.* **9**, 219–222 (1999).
- Minowada, G. *et al.* Vertebrate Sprouty genes are induced by FGF signaling and can cause chondrodysplasia when overexpressed. *Development* **126**, 4465–4475 (1999).
- Reich, A., Sapir, A. & Shilo, B. Sprouty is a general inhibitor of receptor tyrosine kinase signaling. *Development* **126**, 4139–4147 (1999).

- Marshall, C. J. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179–185 (1995).
- Bennett, A. M. & Tonks, N. K. Regulation of distinct stages of skeletal muscle differentiation by mitogen-activated protein kinases. *Science* **278**, 1288–1291 (1997).
- Wu, Z. *et al.* p38 and extracellular signal-regulated kinases regulate the myogenic program at multiple steps. *Mol. Cell. Biol.* **20**, 3951–3964 (2000).
- Gille, H. *et al.* ERK phosphorylation potentiates Elk-1-mediated ternary complex formation and transactivation. *EMBO J.* **14**, 951–962 (1995).
- Yeung, K. *et al.* Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature* **401**, 173–177 (1999).
- Mason, C. S. *et al.* Serine and tyrosine phosphorylations cooperate in Raf, but not B-Raf activation. *EMBO J.* **18**, 2137–2148 (1999).
- King, A. J. *et al.* The protein kinase Pak3 positively regulates Raf-1 activity through phosphorylation of serine 338. *Nature* **396**, 180–183 (1998).
- Ridley, A. J. *et al.* The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell* **70**, 401–410 (1992).
- DeMille, M. M., Kimmel, B. E. & Rubin, G. M. A *Drosophila* gene regulated by *rough* and *glass* shows similarity to ena and VASP. *Gene* **183**, 103–108 (1996).
- Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M. & Davis, R. W. Lambda YES: A multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *Escherichia coli* mutations. *Proc. Natl. Acad. Sci. USA* **88**, 1731–1735 (1991).
- Yokouchi, M. *et al.* Cloning and characterization of APS, an adaptor molecule containing PH and SH2 domains that is tyrosine phosphorylated upon B-cell receptor stimulation. *Oncogene* **15**, 7–15 (1997).
- Yasukawa, H. *et al.* The JAK-binding protein JAB inhibits Janus tyrosine kinase activity through binding in the activation loop. *EMBO J.* **18**, 1309–1320 (1999).
- Nosaka, T. *et al.* STAT5 as a molecular regulator of proliferation, differentiation and apoptosis in hematopoietic cells. *EMBO J.* **18**, 4754–4765 (1999).
- Rose, D. W., McCabe, G., Feramisco, J. R. & Adler, M. Expression of c-fos and AP-1 activity in senescent human fibroblasts is not sufficient for DNA synthesis. *J. Cell Biol.* **119**, 1405–1411 (1992).
- Suzuki, A. *et al.* CIS3/SOCS3/SSI3 plays a negative regulatory role in STAT3 activation and intestinal inflammation. *J. Exp. Med.* **193**, 471–481 (2001).
- Warne, P. H., Viciani, P. R. & Downward, J. Direct interaction of Ras and the amino-terminal region of Raf *in vitro*. *Nature* **364**, 352–355 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

The first two authors contributed almost equally to this work. We thank H. Ohgusu, M. Sasaki and N. Tanaka for technical assistance; T. Sato and Y. Kaziro for wild-type and mutant Ras; H. Miki for the *Drosophila* cDNA library; Y. Hiromi for *Drosophila* Sprouty cDNA; P. Chambon, R. Denton, M. Yokouchi and J.-M. Garnier for library construction; Y. Sako and S. Minoguchi for valuable discussions; and M. Ohara for critical comments on the manuscript and for language assistance. This work was supported in part by grants from the Ministry of Education, Science, Technology, Sports, and Culture of Japan, Japan Research Foundation for Clinical Pharmacology, Fukuoka Cancer Society, Uehara Memorial Foundation and the Mitsubishi Foundation. Murine Spred-1 and Spred-2 are deposited in the nucleotide sequence databases under accession numbers AB063495 and AB063496, respectively.

Correspondence and requests for materials should be addressed to A.Y. (e-mail: yakihiko@bioreg.kyushu-u.ac.jp).

General transcription factors bind promoters repressed by Polycomb group proteins

Achim Breiling*, Bryan M. Turner†, Marco E. Bianchi* & Valerio Orlando*

*DIBIT, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy
†Chromatin and Gene Expression Group, University of Birmingham Medical School, Birmingham B15 2TT, UK

To maintain cell identity during development and differentiation, mechanisms of cellular memory have evolved that preserve transcription patterns in an epigenetic manner. The proteins of the Polycomb group (PcG) are part of such a mechanism, maintaining gene silencing. They act as repressive multiprotein complexes that may render target genes inaccessible to the transcriptional machinery^{1,2}, inhibit chromatin remodelling^{3,4}, influence chromosome domain topology⁵ and recruit histone deacetylases (HDACs)⁶. PcG proteins have also been found to

bind to core promoter regions⁷, but the mechanism by which they regulate transcription remains unknown. To address this, we used formaldehyde-crosslinked chromatin immunoprecipitation (X-ChIP) to map TATA-binding protein (TBP), transcription initiation factor IIB (TFIIB) and IIF (TFIIF), and dHDAC1 (RPD3) across several *Drosophila* promoter regions. Here we show that binding of PcG proteins to repressed promoters does not exclude general transcription factors (GTFs) and that depletion of PcG proteins by double-stranded RNA interference leads to de-repression of developmentally regulated genes. We further show that PcG proteins interact *in vitro* with GTFs. We suggest that PcG complexes maintain silencing by inhibiting GTF-mediated activation of transcription.

For X-ChIP analysis of promoter regions, we chose the PcG target genes *Abdominal-B* (*Abd-B*, B-promoter), *iab-4*, *abdominal-A* (*abd-A*, AI-promoter) and *Ultrabithorax* (*Ubx*), all located in the Bithorax complex (BX-C), *engrailed* (*en*) and *empty spiracles* (*ems*)⁷⁻⁹. In addition, we chose *RpII140* (the subunit of RNA

polymerase II with relative molecular mass 140,000 (M_r 140K)) and *brown* (*bw*), genes that do not reside in PC binding sites on polytene chromosomes¹⁰ and thus are most probably not PcG regulated. We analysed the expression of these genes in *Drosophila* SL-2 culture cells by polymerase chain reaction with reverse transcription (RT-PCR) and found that *Abd-B* and *RpII140* are transcribed whereas *iab-4*, *abd-A*, *Ubx*, *en*, *ems* and *bw* are inactive (Fig. 1).

Acetylation of histones H3 and H4 is considered to be a mark for ongoing transcription. Thus, we screened the promoters of the genes listed above for the presence of amino-terminally acetylated H4 and H3 by X-ChIP (Fig. 1). Two antisera were used, one that recognizes H4 acetylated at lysine 12 and one or more other lysines, and one that recognises H3 acetylated at lysines 9 and/or 18 (ref. 11). H4 was found generally acetylated across the promoter regions analysed, in some cases with reduced levels in upstream and downstream regions (Fig. 1a, c, d, f–h). H3 was strongly acetylated in the active *Abd-B* and *RpII140* promoters (Fig. 1a, g; p10, p10b,

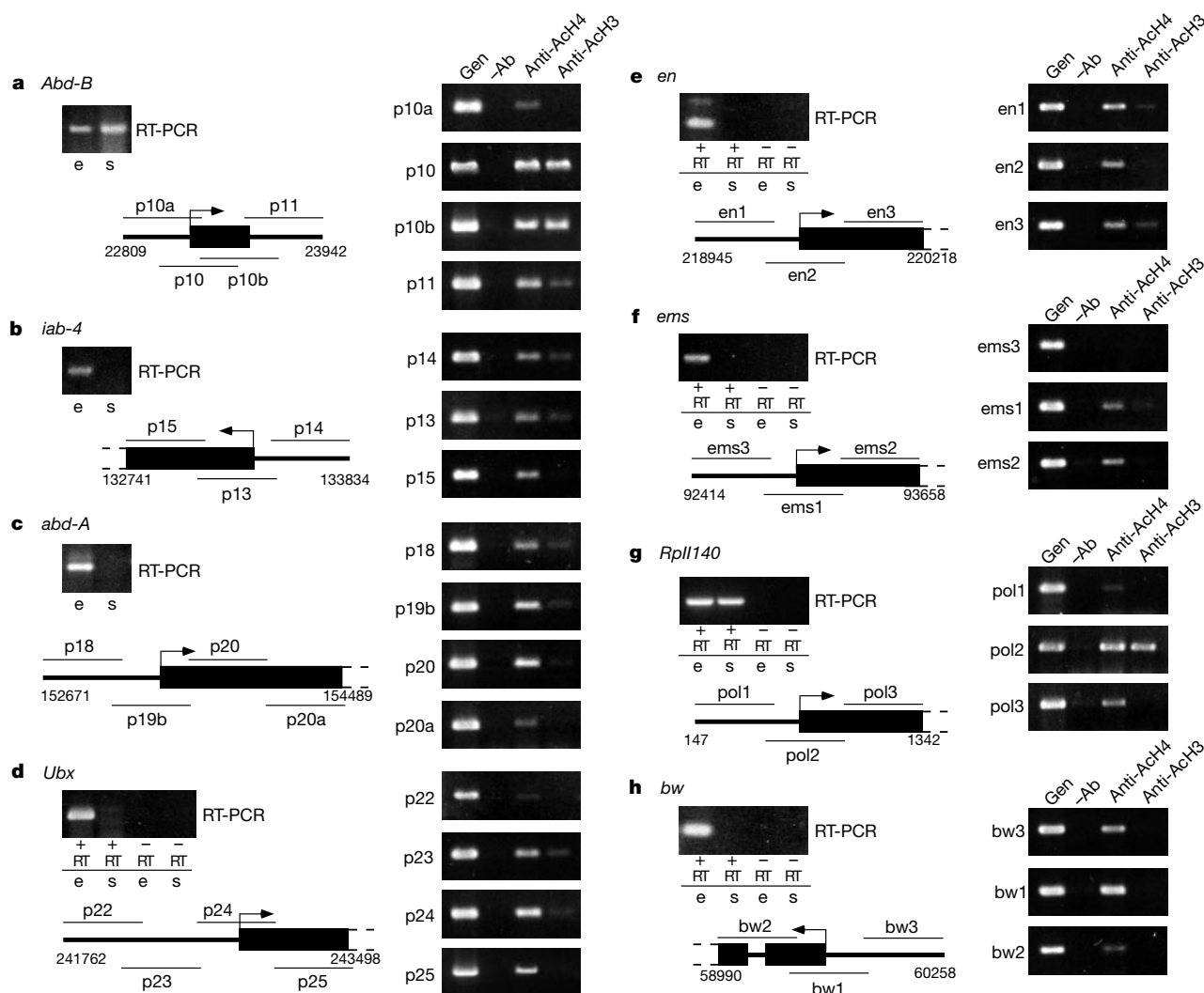


Figure 1 Analysis of the gene expression and acetylation state of BX-C, *engrailed*, *empty spiracles*, *RpII140* and *brown* genes. Left panels show RT-PCR analysis using total RNA from embryos (positive control, e) and SL-2 cells (s) of the promoter indicated. Diagrams show the positions of the primer pairs used and the first exon (black box). Right panels show the PCR analysis of the DNA immunoprecipitated with antibodies specific for acetylated H4 (Anti-Ach4) and acetylated H3 (Anti-Ach3). Genomic DNA served as a positive control (Gen); the addition of no antibody as a negative control (-Ab). Results shown are from one representative experiment. **a**, *Abd-B* promoter. A primer pair in the

first exon of B-class transcripts and in the first exon common to all transcripts was used for the RT-PCR. **b**, *iab-4* promoter. A primer pair in the first and second exons was used for the RT-PCR. **c**, *abd-A* AI promoter. A primer pair in the first and fourth exons was used for the RT-PCR. **d**, *Ubx* promoter. PCR reactions without the RT step are shown to monitor genomic DNA contamination (RT-PCR primer p25). **e**, *en* promoter (RT-PCR primer en3). **f**, *ems* promoter (RT-PCR primer ems2). **g**, *RpII140* promoter (RT-PCR primer pol3). **h**, *bw* promoter (RT-PCR primer bw2).

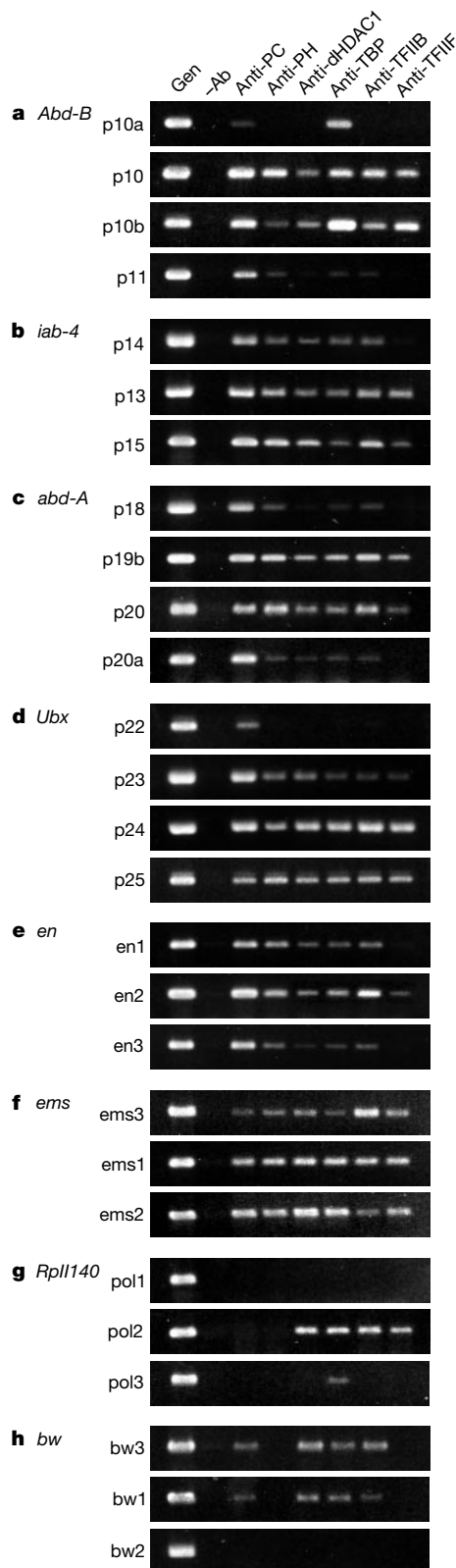


Figure 2 PCR analysis of the protein distribution at BX-C, *engrailed*, *empty spiracles*, *Rpl140* and *brown* promoter regions. Antibodies used for the immunoprecipitation are indicated on top of the panels. A chromatin sample without the addition of antibody was used as negative control (-Ab) and genomic DNA as positive control for the PCR reaction (Gen). **a**, *Abd-B*-promoter. **b**, *iab-4* promoter. **c**, *abd-A* promoter. **d**, *Ubx* promoter. **e**, *en* promoter. **f**, *ems* promoter. **g**, *Rpl140* promoter. **h**, *bw* promoter. Panels shown are from one representative experiment.

pol2), whereas the inactive loci (*iab-4*, *abd-A*, *Ubx*, *en*, *ems* and *bw*) showed a decrease (5–10 times less than the H3 signal in the active *Abd-B* and *Rpl140* promoters) or absence of acetylation both at the core promoters as well as downstream of the initiator (Fig. 1b–f, h). Thus, H3 is acetylated in the active but underacetylated in the inactive promoters, whereas H4 acetylation shows no such changes. Acetylation of histones H3 and H4 seems to be regulated independently across the BX-C, consistent with results in other systems¹².

We analysed the same promoter regions by X-ChIP using antibodies against the PcG proteins Polycomb (PC) and Polyhomeotic (PH), dHDAC1, TBP, TFIIB and TFIIF (RAP 30 subunit, associated with RNA polymerase II). We found all six proteins in the core promoter regions (200 base pairs (bp) around the initiator) of the *Abd-B* (p10), *iab-4* (p13), *abd-A* (p19b), *Ubx* (p24), *en* (en2) and *ems* (ems1) transcription units (Fig. 2a–f). PC was found in most regions both upstream and downstream of the transcription start site, a characteristic spreading that has been reported before^{7–9}.

The presence of PC and PH at the active *Abd-B* promoter is striking, although preceded¹³. Co-localization of PcG proteins and trithorax-group (trxG) proteins, which have been identified as suppressors of the PcG², has been reported for most PcG-bound regulatory regions of the BX-C, including promoters^{7,8,14}. PcG and certain trxG proteins might simultaneously be needed for changing and maintaining opposite transcriptional states. Thus, coincidental association of repressors and activators with active genes might act to guarantee regulated levels of transcription.

Rpl140 was expected to be active and not PcG controlled, and indeed we found GTFs and also dHDAC1, but not PcG proteins, bound to the core promoter (Fig. 2g; pol2). As a putative non-PcG-repressed target, we chose the *bw* gene. Again, we found TBP, TFIIB and dHDAC1 but not TFIIF (Fig. 2h). PC, but not PH, binds to the *bw* promoter. However, we found no PC downstream of the *bw* promoter, unlike the other PcG-controlled genes. This and the absence of PH suggest that *brown* is not a canonical PcG-controlled gene, and PC here might have an as-yet-unknown function, not requiring other PcG proteins.

We could not correlate the acetylation state of the promoters investigated with the presence of dHDAC1. dHDAC1 might be present at active promoters together with histone acetyl transferases (HATs) to maintain an appropriate steady-state level of histone acetylation. Five *Drosophila* HDACs have been identified so far—dHDAC1 (also known as RPD3), dHDAC3, dHDAC4, dHDAC6 and dSIR2 (ref. 15)—and any of these might contribute to the

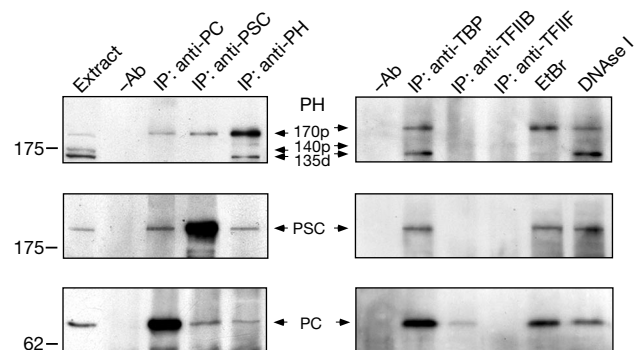


Figure 3 Interaction of TBP with PcG proteins. Western analysis of immunoprecipitations of nuclear extract from *Drosophila* embryos. The antibodies used for the precipitations (IP) are indicated above the lanes (-Ab is the negative control, lacking antibody). An aliquot of the nuclear extract was loaded as input control. TBP immunoprecipitates were treated with ethidium bromide (EtBr) or DNase I (DNase I) to exclude indirect co-purification due to presence of DNA. The western blots were stained using antibodies against PH, PSC and PC. The sizes of PC, PSC and the three PH isoforms (*M*_r 170K, 140K and 135K) are indicated. The marker on the left is prestained protein marker (New England Biolabs), broad range relative molecular mass of 62 and 175 K respectively.

acetylation patterns observed. dHDAC1 might also regulate transcriptional activity by deacetylating transcription factors. GTFs, in particular TFIIE and TFIIIF, are substrates for histone acetyltransferases¹⁶, and transcriptional regulation involves acetylation of developmentally controlled transcription factors¹⁷. The presence of dHDAC1 at promoters could also be due to the association with Topoisomerase II (refs 18,19), which we previously found to be bound to PcG binding sites⁷.

Components of the transcription initiation complex (TBP, TFIIB and TFIIIF) were mapped not only to core promoters but also to regions immediately downstream or upstream of the initiator, but not more than 1 kilobase (kb) away from the transcription start site (Fig. 2; p11, p18, p20a, p22). As the average size of the immunoprecipitated chromatin in these assays is around 0.5–1 kb, the resolution of the X-ChIP approach might be lower than the size of the amplified PCR fragments. This may account for some of the apparent spreading of binding sites.

PcG function is generally thought to be incompatible with binding of GTFs^{1,2}. Thus, the presence of GTFs at PcG-repressed promoters is surprising. TBP binding and the assembly of TFIID is considered to be strictly connected to gene activation and ongoing transcription²⁰. However, inducible promoters have been described in yeast in which TBP occupancy was very significant also before transcriptional activation and only modestly increased after induction, although in a majority of the promoter regions studied a perfect correlation between levels of transcription and TBP

recruitment has been reported²¹. Our finding of TBP bound to inactive as well as active loci is consistent with immunostaining of *Drosophila* polytene chromosomes, showing a rather ubiquitous pattern for TBP²². In the same tissue, discrete patterns of elongating PolII and putatively inactive PolII have been reported²³, suggesting that PolII might be associated with chromatin irrespective of transcription. Our results support the suggestion that this configuration, typical of heat shock gene promoters²⁴, may be more common than currently believed.

The close mapping to the same DNA regions suggested a physical interaction of PcG proteins with TBP and other GTFs. Using antibodies against TBP we showed that the PcG proteins PC, PH and PSC are co-precipitated in a DNA-independent manner with TBP from embryonic nuclear extracts (Fig. 3) and from nuclear extracts from SL-2 cells (not shown). The long proximal isoform of PH (PH 170p) has been found to interact with other PcG proteins^{9,25} and we also found this protein co-purified with PC, PSC and TBP (Fig. 3). Only PC, not PH or PSC, is co-precipitated with TFIIB, whereas we observed no co-purification of PcG proteins with TFIIIF.

PcG proteins at repressed promoters may prevent activation of RNA Polymerase II, otherwise committed to transcribe. We tested this hypothesis by dsRNA interference (RNAi)^{26,27}, a targeted destruction of messenger RNA, to see whether the inhibition of PcG protein synthesis would lead to de-repression of inactive genes. As shown in Fig. 4a, after prolonged treatment of SL-2 cells with *Pc* and *ph* dsRNAs, PC protein was no longer detectable in cellular extracts by western blotting, and the amount of PH was significantly lower than in non-treated cells. With the same kinetics, PcG-regulated promoters, which are inactive in non-treated cells (*iab-4*, *abd-A*, *Ubx*, *en* and *ems*), became de-repressed upon treatment with *Pc* or *ph* dsRNAs (Fig. 4b). In contrast, *bw* did not show any change of expression state, like *Abd-B* and *RpII140*, underlining the specificity of PcG control (Fig. 4b). The dsRNA-treated cells were fully viable and continued to grow for at least 8 days. PcG^{-/-} cell clones in *Drosophila* imaginal discs also continue to grow and show a significant de-repression of homeotic genes²⁸. Remarkably, an incubation time of 8 days was necessary to observe a significant de-repression of PcG target genes. It appears that PcG proteins are rather stable and cells have to divide several times (eight times assuming a duplication time of roughly 24 h) with inhibited PcG protein synthesis, before an effect on transcription is seen.

The major conclusion from this work is that promoters constitute a key target of PcG function. We provide evidence that, unexpectedly, GTFs are retained at PcG-repressed promoters and that PcG proteins may function through direct physical interactions with GTFs. This mechanism of transcriptional regulation may provide both transcriptional competence and the flexibility necessary for the rapid re-arrangement of patterns of gene expression in response to developmental signals. Thus, the presence of GTFs and some trxG proteins^{7,8} at PcG-repressed promoters would allow a relatively fast re-activation of these genes, as differentiation processes require. In this context, PcG proteins would need to be continuously present at target gene promoters to constitutively inhibit transcription, a prediction supported by our finding that PcG-repressed genes are re-expressed in cells depleted of PcG proteins by dsRNA interference.

Methods

Cellular extracts and immunoprecipitation

Nuclear extract preparation and immunoprecipitation were performed as described⁵.

Antibodies

PSC and PH antibodies were produced in rabbits with the epitopes as described⁹ and affinity purified. Antibodies against PC were obtained from R. Paro; against dTBP, dTFIIB and dTFIIIF from J. Kadonaga. Rabbit polyclonal antibodies against acetylated histones H4, H3 and the carboxy-terminal peptide CSGSGTSGAGAKGAKENNI of dHDAC1 were produced as described¹¹.

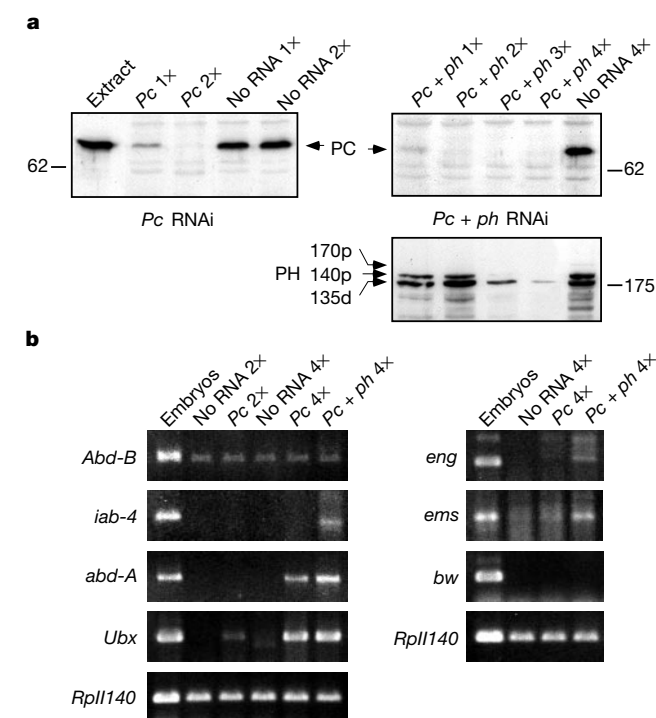


Figure 4 Post-transcriptional silencing of *Pc* and *ph* in SL-2 cells. **a**, Inhibition of PC and PH production by RNAi. Left panel shows a western analysis of total extract of SL-2 cells treated with *Pc* dsRNA after one transfection (*Pc* 1x) and two transfections (*Pc* 2x), non-treated (No RNA 1x and 2x). An aliquot of embryonic nuclear extract was included for size comparison. Right panels show an analysis of cells treated with *Pc* and *ph* dsRNA at the same time, after one (*Pc* + *ph* 1x), two (*Pc* + *ph* 2x), three (*Pc* + *ph* 3x) and four (*Pc* + *ph* 4x) rounds of transfection. Non-treated cells grown for 8 d are shown as negative control (No RNA 4x). The upper blots were probed with PC antibodies, the lower blot with PH antibodies. **b**, RT-PCR analysis of SL-2 cells treated with *Pc* and *ph* dsRNA, using total RNA from embryos (positive control) and RNA prepared from SL-2 cells transfected twice with *Pc* dsRNA (*Pc* 2x), four times with *Pc* dsRNA (*Pc* 4x), four times with *Pc* and *ph* dsRNA (*Pc* + *ph* 4x) and from the respective control cells, grown without the addition of dsRNA (No RNA 2x, No RNA 4x). Marker as in Fig. 3.

X-ChIP and PCR

Crosslinked chromatin was prepared from the cell line SL-2 (grown in serum-free medium; HyQ-CCM 3, Hyclone) and immunoprecipitation was performed as described previously⁷. The precipitated DNA was re-dissolved in 180 µl TE (10 mM Tris buffer at pH 8; 1 mM EDTA) and stored at 4 °C or used directly for PCR. Primer pairs (melting temperature 64–68 °C) amplifying 400–500-bp fragments in the promoter regions of the *BX-C*, *engrailed*, *empty spiracles*, *RpII140* and *brown* were created using the published sequences (NCBI accession numbers: *BX-C*, U31961; *RpII140*, X05709; *engrailed*, AE003825 and M10017; *empty spiracles*, AE0003702 and X66270; and *brown*, AE003461 and L05635). For each primer pair the optimal magnesium concentration (1–2 mM MgCl₂) was determined (Taq polymerase, Promega). PCR scheme: 94 °C for 3 min, once; 94 °C for 1 min, 60 °C for 1 min, 72 °C for 1 min, 34 times; 94 °C for 1 min, 60 °C for 1 min, 72 °C for 7 min, once. PCR with the immunoprecipitated material was performed with the optimal magnesium concentration for each primer pair using 2–3 µl of the template with the same PCR scheme. For individual primer pairs the annealing temperature and number of cycles were adjusted until no signal was detected for the mock-immunoprecipitate DNA, but the amplification on the genomic template (200 ng) was not altered. Signals obtained with the antibody-immunoprecipitate DNA under these conditions were considered significant. The amplified DNA was separated on 1.5% agarose gels and visualized with ethidium bromide.

RT-PCR

Reverse transcription and PCR amplification were performed with the Access RT-PCR System (Promega) using DNase-treated (DNase RNase free, Roche) total RNA from embryos (0–20 h overnight egg lays) or SL-2 culture cells (prepared with the SV Total RNA Isolation System, Promega) as a template (1–2 µg per reaction). The amplified DNA was separated on 1.5% agarose gels and visualized with ethidium bromide.

RNAi

Templates for producing complementary transcripts of ~1,400 bp, corresponding to exonic regions of *Pc* and *ph*, were amplified from genomic DNA using primer pairs incorporating a T7 promoter. One template for antisense and one template for sense strand transcription were obtained. dsRNA production, annealing and RNAi were performed basically as described²⁷. Confluent SL-2 cells were diluted to 0.7 × 10⁶ cells ml⁻¹ into 5 ml of serum-free growth medium (HyQ-CCM 3, Hyclone) in a small culture bottle. 6–8 µl of FuGENE6 transfection reagent (Roche) were mixed with 2–7 µg dsRNA (*Pc*-dsRNA or a mix of *Pc*- and *ph*-dsRNA), incubated for 30 min at room temperature and added to the cells. As a negative control the same number of cells was grown in parallel without the addition of dsRNA. After 2 d of growth, cells (dsRNA-treated and non-treated) were diluted to 0.7 × 10⁶ cells ml⁻¹ in the same flask and the same amount of FuGENE6/dsRNA was added. Up to four sequential transfections were performed (8 d of growth). For western analysis of treated and non-treated cells, 0.5 × 10⁶ cells were pelleted and directly dissolved into SDS–PAGE loading buffer at 95 °C on a shaker, separated on 8% SDS–PAGE minigels and transferred to Immobilon P membranes (Millipore). Proteins were detected with ECL (Amersham).

Received 23 March; accepted 29 June 2001.

- Pirrotta, V. PcG complexes and chromatin silencing. *Curr. Opin. Genet. Dev.* **7**, 249–258 (1997).
- Paro, R. & Harte, P. J. in *Epigenetic Mechanisms of Gene Regulation* (eds Russo, V. E. A., Martienssen, R. A. & Riggs, A. R.) 507–528 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1996).
- Breiling, A., Bonte, E., Ferrari, S., Becker, P. B. & Paro, R. The *Drosophila* Polycomb protein interacts with nucleosomal core particles *in vitro* via its repression domain. *Mol. Cell. Biol.* **19**, 8451–8460 (1999).
- Shao, Z. *et al.* Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37–46 (1999).
- Lupo, R., Breiling, A., Bianchi, M. E. & Orlando, V. *Drosophila* chromosome condensation proteins Topoisomerase II and Barren colocalize with Polycomb and maintain Fab-7 PRE silencing. *Mol. Cell* **7**, 127–136 (2001).
- van der Vlag, J. & Otte, A. P. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nature Genet.* **23**, 474–478 (1999).
- Orlando, V., Jane, E. P., Chinwalla, V., Harte, P. J. & Paro, R. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* **17**, 5141–5150 (1998).
- Strutt, H., Cavalli, G. & Paro, R. Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.* **16**, 3621–3632 (1997).
- Strutt, H. & Paro, R. The Polycomb Group protein complex of *Drosophila melanogaster* has different compositions at different target genes. *Mol. Cell. Biol.* **17**, 6773–6783 (1997).
- Paro, R. & Zink, B. The Polycomb gene is differentially regulated during oogenesis and embryogenesis of *Drosophila melanogaster*. *Mech. Dev.* **40**, 37–46 (1992).
- White, D. A., Belyaev, N. D. & Turner, B. M. Preparation of site-specific antibodies to acetylated histones. *Methods* **19**, 417–424 (1999).
- Gregory, R. I. *et al.* DNA methylation is linked to deacetylation of histone H3, but not H4, on the imprinted genes *Snrpn* and *U2af1-rs1*. *Mol. Cell. Biol.* (in the press).
- Cavalli, G. & Paro, R. Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* **286**, 955–958 (1999).
- Chinwalla, V., Jane, E. P. & Harte, P. J. The *Drosophila* trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. *EMBO J.* **14**, 2056–2065 (1995).
- Barlow, A. L. *et al.* dSIR2 and HDAC6; two novel, inhibitor-resistant deacetylases in *Drosophila melanogaster*. *Exp. Cell Res.* **265**, 90–103 (2001).
- Imhof, A. *et al.* Acetylation of general transcription factors by histone acetyltransferases. *Curr. Biol.* **7**, 689–692 (1997).

- Waltzer, L. & Bienz, M. *Drosophila* CBP represses the transcription factor TCF to antagonize Wingless signalling. *Nature* **395**, 521–525 (1998).
- Tsai, S. C. *et al.* Histone deacetylase interacts directly with DNA topoisomerase II. *Nature Genet.* **26**, 349–353 (2000).
- Johnson, C. A., Padgett, K., Austin, C. A. & Turner, B. M. Deacetylase activity associates with topoisomerase II and is necessary for etoposide-induced apoptosis. *J. Biol. Chem.* **276**, 4539–4542 (2001).
- Ptashe, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
- Kuras, L. & Struhl, K. Binding of TBP to promoters *in vivo* is stimulated by activators and requires Pol II holoenzyme. *Nature* **399**, 609–613 (1999).
- Hansen, S. K., Takada, S., Jacobson, R. H., Lis, J. T. & Tjian, R. Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell* **91**, 71–83 (1997).
- Weeks, J. R., Hardin, S. E., Shen, J., Lee, J. M. & Greenleaf, A. L. Locus-specific variation in phosphorylation state of RNA polymerase II *in vivo*: correlations with gene activity and transcript processing. *Genes Dev.* **7**, 2329–2344 (1993).
- Rougvié, A. E. & Lis, J. T. The RNA polymerase II molecule at 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**, 795–804 (1988).
- Kyba, M. & Brock, H. W. The *Drosophila* Polycomb Group protein Psc contacts Ph and Pc through specific conserved domains. *Mol. Cell. Biol.* **18**, 2712–2720 (1998).
- Hammond, S. M., Caudy, A. A. & Hannon, G. J. Post-transcriptional gene silencing by double-stranded RNA. *Nature Rev. Genet.* **2**, 110–119 (2001).
- Wei, Q., Marchler, G., Edington, K., Karsch-Mizrachi, I. & Paterson, B. M. RNA interference demonstrates a role for nautilus in the myogenic conversion of Schneider cells by daughterless. *Dev. Biol.* **228**, 239–255 (2000).
- Beuchle, D., Struhl, G. & Muller, J. Polycomb group proteins and heritable silencing of *Drosophila* Hox genes. *Development* **128**, 993–1004 (2001).

Acknowledgements

We thank B. M. Paterson for advice concerning the RNAi experiments and J. Kadonaga for discussions and comments on the manuscript. We are indebted to J. Butler, J. Kadonaga, A. Barlow and R. Paro for providing (respectively) dTBP, dTFIIB, dTFIIF, dHDAC1 and PC antibodies. We thank N. Collu for technical assistance. This work was supported by a postdoctoral fellowship to A.B. by EU-TMR, and by research grants from A.I.R.C. and TELETHON to V.O., from the Wellcome Trust and the Human Frontier Science Program to B.M.T. and from MURST to M.E.B.

Correspondence and requests for materials should be addressed to V.O. (e-mail: v.orlando@hsr.it).

A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins

Andrew J. Saurin*, Zhaohui Shao*†, Hediye Erdjument-Bromage‡, Paul Tempst‡ & Robert E. Kingston*

* Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

A goal of modern biology is to identify the physical interactions that define ‘functional modules’¹ of proteins that govern biological processes. One essential regulatory process is the maintenance of master regulatory genes, such as homeotic genes, in an appropriate ‘on’ or ‘off’ state for the lifetime of an organism. The Polycomb group (PcG) of genes maintain a repressed transcriptional state, and PcG proteins form large multiprotein complexes^{2,3}, but these complexes have not been described owing to inherent difficulties in purification. We previously fractionated a major PcG complex, PRC1, to 20–50% homogeneity from *Drosophila* embryos. Here, we identify 30 proteins in these preparations, then further fractionate the preparation and use western analyses to validate unanticipated connections. We

† Present address: Biogen, Inc., 14 Cambridge Center, Cambridge, Massachusetts 02142, USA.

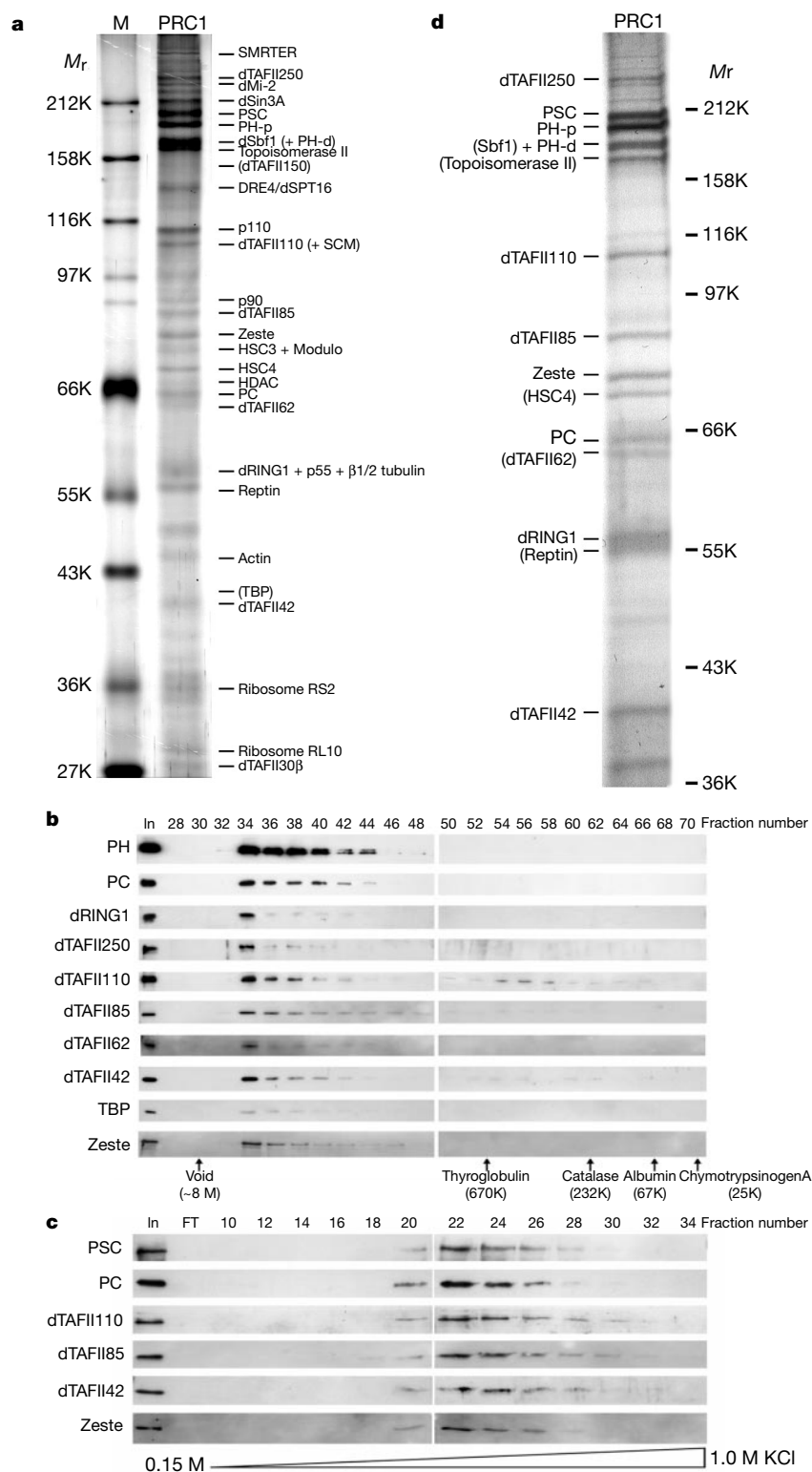


Figure 1 Identification and analysis of proteins identified in PRC1 preparations. **a**, Silver stain of a preparative 8% SDS-PAGE-resolved PRC1 preparation used to identify gel-isolated protein species by mass spectrometry analysis. Proteins identified only by western blot analyses are given in parentheses. Relative molecular mass of standards (M_r , in thousands (K)) are indicated. **b**, Western blot analysis of the fractionation profile of PcG proteins, dTAFII proteins, TBP and Zeste following gel filtration of PRC1. Lanes correspond to 25 μ l of even-numbered fractions. In, ~12 fmol M2-purified PRC1. Peak elutions of standards during gel filtration are indicated. **c**, Western blot analysis of the elution profiles of PSC, PC, dTAFII110, dTAFII85, dTAFII42 and Zeste during fractionation of PRC1 over a

heparin affinity column. In, ~12 fmol M2-purified PRC1; FT, 8 μ l flow-through. Fractions represent 40 μ l (PSC, PC, dTAFII110, dTAFII85 and dTAFII42) or 80 μ l (Zeste) of fractions eluting during the KCl gradient. The peak elution of the proteins shown occur at ~450 mM KCl. **d**, Silver stain of a 7.5% SDS-PAGE-resolved peak PRC1 fraction following heparin fractionation. Proteins unconfirmed by western analyses are shown in parentheses. Note that PSC and PH both stain preferentially with silver, and that most of Topoisomerase II fractionates away from PRC1 during gel filtration chromatography (see Supplementary Information, and data not shown).

show that the known PcG proteins Polycomb, Posterior sex combs, Polyhomeotic and dRING1 exist in robust association with the sequence-specific DNA-binding factor Zeste and with numerous TBP (TATA-binding-protein)-associated factors that are components of general transcription factor TFIID (dTAFIIs). Thus, in fly embryos, there is a direct physical connection between proteins that bind to specific regulatory sequences, PcG proteins, and proteins of the general transcription machinery.

The inheritance of established expression patterns of certain genes during multiple cell divisions is essential for the correct development of an animal. In *Drosophila melanogaster*, the expression patterns of the homeotic genes that govern body segment identity are established early in embryogenesis by the products of the gap and pair-rule genes, but are maintained throughout the rest of development by proteins of the PcG and trithorax group (trxG) (ref. 4 and references therein). The trxG maintains the transcriptionally active state of the homeotic genes, whereas the PcG prevents ectopic expression by maintaining a repressive state. The PcG genes encode components of multiple complexes. One of these complexes, Polycomb repressive complex 1 (PRC1), contains the Polycomb (PC), Polyhomeotic (PH) and Posterior sex combs (PSC) proteins⁵.

To better understand the mechanisms of this cellular memory system, we previously used an epitope-tag strategy to purify PRC1 over 3,000-fold from *Drosophila* embryos⁵. This complex has been extensively washed in 1 M salt and has a high specific activity in functional analyses; however, contaminating proteins remain associated. Extensive efforts to fractionate this complex to homogeneity in reasonable quantity were blocked by unacceptably low yields on a wide variety of subsequent purification steps. The advent of genome-wide sequence analysis provided an alternative route to identify the components of PRC1. Using mass spectrometry and the recently completed *Drosophila* genome, we identified almost all of the proteins in the highly fractionated material derived from the M2-affinity column (Fig. 1a). We then used the sensitivity of western analysis to validate the association of proteins during subsequent chromatography steps (Fig. 1b, c).

The presence of the previously identified PcG proteins PH, PC, and PSC was confirmed by mass spectrometry (Fig. 1). In addition, a *Drosophila* homologue of the mammalian RING1 protein (dRING1) was identified; dRING1 has been found to colocalize with PC on polytene chromosome preparations (M. A. Vidal and S. Pimpinelli, personal communication) and its mammalian counterparts have previously been shown to associate with mammalian PcG proteins^{6,7}. Thus, the PcG complement of PRC1 is made up from PH, PSC, PC, dRING1 and sub-stoichiometric amounts of Sex Combs on Midleg (SCM).

Of the remaining proteins identified by mass spectrometry, the presence of several dTAFII proteins and Zeste is particularly striking (Fig. 1). Zeste is a sequence-specific DNA-binding factor, with binding sites in the promoter and regulatory regions of some homeotic genes⁸. The dTAFII proteins were initially identified in the general transcription factor TFIID, a central component for transcriptional initiation, but are also found in histone acetyltransferase complexes (for a recent review, see ref. 9).

To validate the unexpected direct association of these proteins with PcG proteins, we further fractionated PRC1 using two independent chromatographic steps. Gel filtration fractionation of PRC1 preparations on Sephacryl S-400 followed by immunoblotting analyses shows that dTAFIIs 250, 110, 85, 62 and 42 and Zeste are tightly associated with the PcG proteins in a very large macromolecular complex (Fig. 1b and see Supplementary Information). Sub-stoichiometric amounts of TBP are also associated. Several proteins fractionated away from PRC1 during gel filtration, demonstrating that these proteins are not tightly associated with PRC1 (Table 1 and see Supplementary Information). In a separate step, we used heparin-agarose chromatography and eluted with a shallow salt gradient to improve resolution. We found that PcG, dTAFII and Zeste proteins all elute in the same fractions (Fig. 1c), further demonstrating their tight association. The peak fractions isolated after heparin-agarose chromatography show less complexity than the input M2-affinity-purified material, demonstrating the efficacy of this step (Fig. 1d).

Table 1 Proteins identified by mass spectrometry from *Drosophila* PRC1 preparations

Protein	Accession number	Predicted M_r ($\times 1,000$)	Stoichiometry relative to PC	Previously linked to PcG?	Previously linked to gene regulation?	Cofractionates with PRC1?
PSC*	P35820	170	0.9	Yes	Yes	Yes*
PH*	P39769	167	1.0	Yes	Yes	Yes*
PC*	P26017	44	1.0	Yes	Yes	Yes*
dRING1*	2388783	47	0.8	Yes	Yes	Yes*
Zeste*	P09956	61	0.8	Yes	Yes	Yes*
HSC4	P11147	71	0.5	Yes	Yes	Yes†
dTAFII250*	P51123	233	0.6	No	Yes	Yes*
dTAFII110*	P47825	100	0.9	No	Yes	Yes*
dTAFII85*	P49846	80	0.9	No	Yes	Yes*
dTAFII62*	740569	64	1.0	No	Yes	Yes*
dTAFII42*	Q27272	29	0.8	No	Yes	Yes*
dTAFII30β	P49906	22	1.3	No	Yes	Unknown‡
SMRTER	5815245	364	0.5	No	Yes	Unknown‡
dMi-2	4325130	224	0.5	Yes	Yes	Unknown‡
dSin3A	2570794	190	ND	No	Yes	Unknown‡
HDAC	7292522	58	0.5	Yes	Yes	Unknown‡
p55	Q24572	48	0.8	Yes	Yes	Unknown‡
dSbf1	3851594	199	0.7	No	Yes	Yes†
DRE4/dSPT16	2511745	123	2.1	No	Yes	Unknown‡
p90	7295314	87	1.9	No	Yes	Unknown‡
HSC3	P29844	72	0.4	No	No	Unknown‡
Modulo	P13469	60	0.4	No	Yes	Unknown‡
Reptin	7293815	53	0.7	No	Yes	Yes†
dTopol*	P15348	164	×	No	Yes	No§
p110	7290971	108	×	No	No	No†
Tubulin	158739	50	×	No	No	Unknown‡
Actin	P10987	42	×	No	No	No§
Ribosome RS2	P31009	29	×	No	No	No§
Ribosome RL10	O61231	25	×	No	No	No§

Stoichiometry was determined by semi-quantitative analysis of the intensity of Coomassie-stained protein bands (see Methods). ×, stoichiometry not given owing to fractionation away from PRC1 during gel filtration; ND, not determined.

* Presence confirmed by western analysis.

† As determined by silver stain analyses.

‡ Presence of multiple proteins or contaminating proteins or insufficient protein precludes identification in peak PRC1 fractions by silver stain analyses.

§ Most of protein fractionates away from peak PRC1 fractions.

By quantifying colloidal Coomassie-stained polyacrylamide gels, we generated a crude estimate of the stoichiometry of the identified proteins in these complexes (Table 1). The dTAII proteins identified by mass spectrometry and Zeste all appear approximately stoichiometric with the PcG complement in the M2 fraction (Table 1), and maintain a quantitative association with PcG proteins on gel filtration and heparin agarose (Fig. 1b–d).

In a separate approach to study the association between dTAII proteins and PcG proteins, we performed immunoprecipitations using antisera specific to various dTAIIs and to TBP (Fig. 2). Using Bio-Rex 70-fractionated embryo extracts, immunoprecipitations with antisera specific to dTAII110, dTAII85 or dTAII42 results in the co-precipitation of PC, PH and PSC (Fig. 2a, lanes 4–12). TBP antisera also co-precipitates PSC, PC and PH, but in lower amounts than that co-precipitated by dTAII antisera (Fig. 2a, lanes 13–15). Additionally, anti-RING1 antisera co-precipitates dTAII110, dTAII85 and dTAII42 (data not shown). PcG proteins co-precipitated with dTAII proteins in the presence of ethidium bromide (Fig. 2b and data not shown), indicating that these associations are unlikely to require DNA. To demonstrate the specificity of these interactions, we found that the PcG protein Enhancer of zeste (E(z)), which resides in a distinct PcG complex¹⁰, does not co-immunoprecipitate with any of the antisera (Fig. 2a).

PRC1 fractionates as an extremely large complex, and contains several other proteins in addition to the PcG, dTAII and Zeste proteins described above. The sequence information (Fig. 1a) provides speculative information on the identity of these proteins, but further work is needed to validate each of these associations. The constitutively expressed Heat shock cognate 3 and 4 (HSC3 and HSC4) proteins were found. The requirement for HSCs in PcG action during development has been demonstrated genetically in flies, where a mutant allele of *HSC4* enhances the homeotic phenotype of *PC*-heterozygous flies¹¹. Proteins were found that have been linked to histone deacetylase complexes, including HDAC (RPD3) (ref. 12), dMi-2 (ref. 13), dSin3A (ref. 14), p55 (ref. 15) and SMRTER, a functional homologue of the human SMRT/N-CoR corepressors¹⁶ (Fig. 1). While dMi-2 has previously been linked genetically to PcG repression¹³, and HDAC and p55 have recently been found present in the Esc/E(z) PcG complex¹⁷, further studies

are clearly needed to examine their association with PRC1. These proteins are present in low stoichiometry (Table 1), cofractionation of these proteins with PcG on subsequent steps could not be accurately assessed owing to lack of signal, and PRC1 has low deacetylase activity when acetylated core histones and histone peptides are used as substrate (data not shown).

The most surprising connection revealed in this study is that between PcG proteins and several dTAIIs. TAI proteins have previously been found in TFIID and in histone acetyltransferase complexes, and in both contexts have been linked to transcriptional activation. This study suggests that they may also function in PRC1-mediated PcG repression. PcG complexes are targeted to specific genes by sequences called Polycomb response elements (PREs), but are also known to associate at promoters¹⁸. The presence of dTAII proteins in PRC1 provides a direct physical connection between PcG proteins and components of the general transcription machinery that bind at promoters. In addition, several of these dTAIIs have similarities with core histone proteins (dTAIIs 62, 42 and 30 β) and have been biochemically and structurally demonstrated to associate with each other in a histone octamer-like substructure^{19–21}. Although quite speculative, the structural similarities between the dTAII42/62 heterotetramer with the histone H3/H4 heterotetramer might indicate a direct role in interacting with nucleosomes and/or DNA to help maintain a stable association of PcG proteins across the numerous rapid cell divisions of the embryo.

The presence of Zeste in PRC1 may serve to assist in the targeting of PcG proteins to repressed loci. Indeed, Zeste can be found localized with PcG proteins at some PcG-repressed loci²² and recent data demonstrates that Zeste is directly involved in the maintenance of the repressed state of some of these loci (M.-W. Hur and M. Biggin, unpublished data). Zeste binds to both PRE and promoter sequences, and thus may serve to bridge the connection of the PcG proteins to these elements. Zeste has also been shown to interact directly with the BRM complex of the *trxG*²³. Zeste thus appears to be involved in both PcG function and *trxG* function, consistent with previous genetic studies implying a role in activation and repression.

Previous work has suggested that the composition of PcG complexes change as organisms develop^{10,18,24}. We have identified

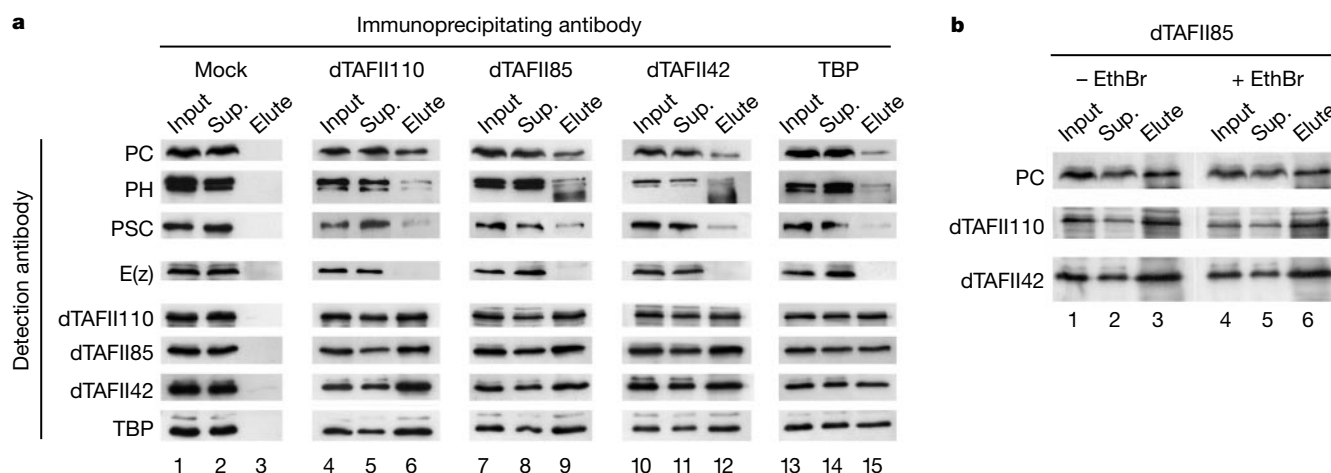


Figure 2 dTAII proteins co-precipitate PRC1 PcG proteins from partially purified *Drosophila* embryo extracts. **a**, Bio-Rex 70-purified FPH embryo extracts were used to immunoprecipitate proteins using specific antisera against dTAII110 (lanes 4–6), dTAII85 (lanes 7–9), dTAII42 (lanes 10–12) or TBP (lanes 13–15). Control immunoprecipitation was performed with normal rabbit sera (lanes 1–3). Co-precipitating proteins were analysed using sera against dTAII110, dTAII85, dTAII42, TBP, PSC, PH, PC or E(z). Lanes represent 1.5 μ l (8.7 μ g) of the precleared Bio-Rex 70-purified embryo

extract (input: lanes 1, 4, 7, 10 and 13), 1.5 μ l of the non-precipitating protein solution (sup.: lanes 2, 5, 8, 11 and 14), and 10 μ l of the immunoprecipitated protein solution (elute: lanes 3, 6, 9, 12 and 15). **b**, Association of dTAII proteins with PcG proteins is not dependent on DNA, as immunoprecipitation of dTAII85 in the absence (lanes 1–3) or presence (lanes 4–6) of ethidium bromide (EthBr) results in the co-precipitation of dTAII110, dTAII42 and PC. Lanes are loaded as in **a**.

proteins associated with PcG proteins in 0–12-h *Drosophila* embryos. We expect that the proteins associated with the core PcG proteins of PRC1 (PSC, PC, PH and dRING1) will change during development. The human homologues of these PcG proteins are also found in a complex in HeLa cells, a cell line isolated from an adult, although there are far fewer proteins associated with this complex (A. Weiss *et al.*, manuscript in preparation). The advances in genome research and proteomics that have allowed the identification of the components of embryonic PRC1 will allow a characterization of how these components change during development. The connection that we have established between Zeste, the PcG proteins and the general transcription machinery provides a means to link PcG complex formation with both PRE sequences and promoter elements. The stable association between these proteins might be particularly important during embryonic development, where structures that maintain gene expression patterns through the life of the organism must be formed in a manner that can be robustly maintained. □

Methods

Mass spectrometry analyses of PRC1 proteins

PRC1 was purified from FPH 0–12-h *Drosophila* embryos as previously described⁵. Approximately 1 pmol of purified PRC1 (6 µg) was resolved by 8% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE) and proteins were visualized by Coomassie blue staining using a Colloidal Blue staining kit (Novex). Stained bands were excised from the gels, digested with trypsin and processed for mass spectrometric fingerprinting as described²⁵. Briefly, peptide mixtures were partially fractionated on Poros 50 R2 RP micro-tips and resulting peptide pools analysed by matrix-assisted laser-desorption/ionization reflectron time-of-flight (MALDI-reTOF) mass spectrometry using a Reflex III instrument (Brüker Franzen), and, in selected cases, by electrospray ionization (ESI) tandem mass spectrometry on an API 300 triple quadrupole instrument (PE-SCIEX), modified with a custom-made fine ionization source, as described²⁶. Selected mass values from the MALDI-TOF experiments were taken to search the Celera/Berkeley *Drosophila* protein database (from NCBI) using the PeptideSearch²⁷ algorithm. Tandem mass spectra from the ESI triple quadrupole analyses were inspected for *y*⁺ ion series and the resultant information transferred to the SequenceTag²⁸ and PepFrag²⁹ programs and used as a search string. Any identification thus obtained was verified by comparing the computer-generated fragment ion series of the predicted tryptic peptide with the experimental tandem mass-spectrometry data.

Estimating stoichiometry of PRC1 proteins

M2-purified PRC1 (~260 fmol) was resolved by 8% SDS–PAGE and visualized by colloidal Coomassie staining using Colloidal Blue stain. The destained gel was digitally imaged with a Kodak 440CF ImageStation (Eastman Kodak). The relative staining intensities of the bands were quantified with ImageQuant (Molecular Dynamics) and expressed as a ratio to the staining intensity obtained for PC. This is a semi-quantitative approach to determining the stoichiometries of proteins present in PRC1, as errors will occur owing to differences in proteins taking up the Coomassie colloid dye, any minor contaminating peptides present, or errors in estimating gel band intensities.

Gel filtration chromatography

PRC1, purified using M2 anti-Flag beads⁵ (Sigma) was further purified by Sephacryl S-400 HR (Amersham Pharmacia) gel filtration chromatography as described³ with minor modifications: 200 µl PRC1 (~8 nM) was loaded onto a Sephacryl S-400 HR column (18.5 ml, 0.7 × 50 cm) in HEGN buffer (25 mM HEPES, K⁺ at pH 7.9, 0.1 mM EDTA, 10% glycerol, 0.1% NP-40, 1 mM dithiothreitol (DTT), 0.1 mM phenylmethyl sulphonyl fluoride (PMSF)), containing 0.3 M KCl (0.3-HEGN) and 50 µg ml⁻¹ insulin (Sigma) and fractionated with a linear flow rate of 4 cm h⁻¹. Subsequent fractions were analysed by 8% SDS–PAGE, followed by immunoblotting with the indicated antibodies. The column was calibrated with a selection of protein markers from gel filtration calibration kits for high and low relative molecular mass (Amersham Pharmacia). Owing to the lack of suitable size markers, the void volume was estimated from the manufacturer's specifications.

Heparin affinity chromatography

PRC1, purified with M2 anti-Flag beads, was diluted to 150 mM KCl with 0-HEGN buffer and applied to a 1-ml HiTrap heparin sepharose high performance column (Amersham Pharmacia) that had been pre-equilibrated with 0.15-HEGN buffer. Following elution of unbound sample with 0.15-HEGN buffer, proteins were eluted with a 20-column volume gradient to 2 M KCl with 2.0-HEGN buffer. Fractions were analysed by 8% SDS–PAGE, followed by immunoblotting with the indicated antibodies. Fractions with salt concentrations up to 1 M KCl are shown, although no additional antibody-specific signal was detected in fractions containing 1–2 M KCl (data not shown). Silver stain analysis was carried out on fractions precipitated with 10% trichloroacetic acid with 0.015% deoxycholic acid added as co-precipitant and resolved on a pre-cast 7.5% polyacrylamide gel (Bio-Rad).

Immunoprecipitation of dTAFII and TBP proteins

Immunoprecipitation of dTAFII proteins was carried out on extracts prepared from FPH 0–12-h embryos partially purified by Bio-Rex 70 chromatography as previously described for the purification of PRC1 (ref. 5). Proteins eluting at 0.85 M KCl were dialysed against 0.3-HEGN buffer and supplemented with 1 µg ml⁻¹ aprotinin, 1 µg ml⁻¹ leupeptin and 50 µg ml⁻¹ TLCK. Extracts were precleared for 1 h at 4 °C using normal rabbit immunoglobulin-γ (Sigma) and protein G Sepharose (Amersham Pharmacia). Immunoprecipitations were then carried out using the appropriate specific antisera for 1 h at 4 °C. Immunocomplexes were isolated by incubation for 30 min with 20 µl protein G Sepharose followed by centrifugation at 1,000 g for 2 min. Following two 500-µl washings of the Sepharose beads with 0.425-HEGN buffer, precipitating proteins were eluted with 80 µl 12.5 mM Tris buffer at pH 2.5 with 0.1 M glycine followed by centrifugation at 2,000 rpm for 2 min. The isolated protein solution was neutralized by the addition of 8 µl 1 M Tris buffer at pH 9.0 and 10 µl was analysed by SDS–PAGE followed by western blotting. For immunoprecipitations in the presence of ethidium bromide, ethidium bromide (Life Technologies) was added to the precleared protein solution at a final concentration of 50 µg ml⁻¹ and incubated for 30 min before immunoprecipitation as described above.

Antibodies

The following antibodies against PRC1 proteins were used during this study: anti-Flag (M5; Sigma), anti-PH and anti-PC (PH21 and PC20; described in ref. 5), anti-Ring1a (a gift of M. Vidal), anti-PSC (6E8 and 7E10, gifts of P. Adler), anti-E(z) (a gift of R. Jones), anti-dTAFIIs 250, 110, 85, 42, 30α and dTBP (gifts of Y. Nakatani), anti-TBP (58C9; Santa Cruz Biotechnology), anti-dTAFIIs 150, 62 and anti-Zeste (gifts of P. Verrijzer), a second anti-Zeste antiserum (a gift of V. Pirrotta), and anti-dTopoisomerase II (a gift of P. Smith).

Received 24 April; accepted 4 July 2001.

- Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
- Franke, A. *et al.* Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* **11**, 2941–2950 (1992).
- Locke, J., Kotarski, M. A. & Tartof, K. D. Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect. *Genetics* **120**, 181–198 (1988).
- Simon, J. Locking in stable states of gene expression: transcriptional control during *Drosophila* development. *Curr. Biol.* **7**, 376–385 (1995).
- Shao, Z. *et al.* Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37–46 (1999).
- Satijn, D. P. E. *et al.* RING1 is associated with the Polycomb-group protein complex and acts as a transcriptional repressor. *Mol. Cell. Biol.* **17**, 4105–4113 (1997).
- Schoorlemmer, J. *et al.* Ring1A is a transcriptional repressor that interacts with the Polycomb-M33 protein and is expressed at rhombomere boundaries in the mouse hindbrain. *EMBO J.* **16**, 5930–5942 (1997).
- Benson, M. & Pirrotta, V. The *Drosophila* zeste protein binds cooperatively to sites in many gene regulatory regions: implications for transvection and gene regulation. *Development* **7**, 3907–3915 (1988).
- Albright, S. R. & Tjian, R. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242**, 1–13 (2000).
- Ng, J., Hart, C. M., Morgan, K. & Simon, J. A. A *Drosophila* ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. *Mol. Cell. Biol.* **20**, 3069–3078 (2000).
- Mollaaghababa, R. *et al.* Mutations in *Drosophila* heat shock cognate 4 are enhancers of Polycomb. *Proc. Natl Acad. Sci. USA* **98**, 3958–3963 (2001).
- De Rubertis, F. *et al.* The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* **384**, 589–591 (1996).
- Kehle, J. *et al.* dMi-2, a Hunchback-interacting protein that functions in Polycomb repression. *Science* **282**, 1897–1900 (1998).
- Pennetta, G. & Pauli, D. The *Drosophila* Sin3 gene encodes a widely distributed transcription factor essential for embryonic viability. *Dev. Genes Evol.* **208**, 531–536 (1998).
- Tyler, J., Bulger, M., Kamakaka, R. T., Kobayashi, R. & Kadonaga, J. T. The p55 subunit of *Drosophila* chromatin assembly factor 1 is homologous to a histone deacetylase-associated protein. *Mol. Cell. Biol.* **16**, 6149–6159 (1996).
- Tsai, C.-C., Kao, H.-Y., Yao, T.-P., McKeown, M. & Evans, R. M. SMRTer, a *Drosophila* nuclear receptor coregulator, reveals that EcR-mediated repression is critical for development. *Mol. Cell* **4**, 175–186 (1999).
- Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E. & Harte, P. J. The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* **128**, 275–286 (2001).
- Orlando, V., Jane, E. P., Chinwalla, V., Harte, P. J. & Paro, R. Binding of Trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* **17**, 5141–5150 (1998).
- Hoffman, A. *et al.* A histone octamer-like structure within TFIID. *Nature* **380**, 356–359 (1996).
- Nakatani, Y., Bagby, S. & Ikura, M. The histone folds in transcription factor TFIID. *J. Biol. Chem.* **271**, 6575–6578 (1996).
- Xie, X. *et al.* Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380**, 316–322 (1996).
- Rastelli, L., Chan, C. S. & Pirrotta, V. Related chromosome binding sites for zeste, suppressors of zeste and the Polycomb group proteins in *Drosophila* and their dependence on Enhancer of zeste function. *EMBO J.* **12**, 1513–1522 (1993).
- Kal, A. J., Mahmoudi, T., Zak, N. B. & Verrijzer, C. P. The *Drosophila* Brahma complex is an essential coactivator for the trithorax group protein Zeste. *Genes Dev.* **14**, 1058–1071 (2000).
- Lessard, J., Baban, S. & Sauvageau, G. Stage-specific expression of polycomb group genes in human bone marrow cells. *Blood* **91**, 1216–1224 (1998).

25. Erdjument-Bromage, H. *et al.* Micro-tip reversed-phase liquid chromatographic extraction of peptide pools for mass spectrometry. *J. Chromatogr.* **826**, 157–181 (1998).
26. Geromanos, S., Freckleton, G. & Tempst, P. Tuning of an electrospray ionization source for maximum peptide-ion transmission into a mass spectrometer. *Anal. Chem.* **72**, 779–790 (2000).
27. Mann, M., Hojrup, P. & Roepstorff, P. Use of mass spectrometric molecular weight information to identify proteins in databases. *Biol. Mass Spectrom.* **22**, 338–345 (1993).
28. Mann, M. & Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390–4399 (1994).
29. Fenyo, D., Qin, J. & Chait, B. T. Protein identification using mass spectrometric information. *Electrophoresis* **19**, 998–1005 (1998).

Supplementary information is available at Nature's World-Wide Web site (<http://www.nature.com>) or as a paper copy from the London editorial office of Nature.

Acknowledgements

We thank A. Nazarian and A. Grewal for help with mass spectrometric analysis, and the numerous people mentioned in 'Methods' for their gifts of antibodies. We also thank N. Francis, K.-M. Lee, S. Levine and A. Weiss for critical reading of the manuscript and E. Duprez and members of the Kingston lab for discussions and comments. A.J.S. is a Human Frontier Science Program Fellow. This work was supported by a NCI Cancer Center grant to P.T. and a NIH grant to R.E.K.

Correspondence and requests for materials should be addressed to R.E.K. (e-mail: kingston@frodo.mgh.harvard.edu).

correction

A β peptide vaccination prevents memory loss in an animal model of Alzheimer's disease

Dave Morgan, David M. Diamond, Paul E. Gottschall, Kenneth E. Ugen, Chad Dickey, John Hardy, Karen Duff, Paul Jantzen, Giovanni DiCarlo, Donna Wilcock, Karen Connor, Jaime Hatcher, Caroline Hope, Marcia Gordon & Gary W. Arendash

Nature **408**, 982–985 (2000).

The caption to Fig. 1 should have included the following statement: "Two transgenic mice vaccinated with A β consistently failed to make choices in the radial arm water maze during the 15.5 month testing period and could not be included in the statistical analysis. Thus the sample size is smaller for the behavioural studies performed at 15.5 mo than at 11.5 mo."

All statistics, results (including means and standard errors in the figures), and the degrees of freedom and significance values in the text are correct as published. We have since replicated our findings with larger numbers of A β vaccinated transgenic mice ($n = 20$)

whose performance in the radial arm water maze is again better than that of non-vaccinated transgenic mice. □

errata

LTRPC7 is a Mg-ATP-regulated divalent cation channel required for cell viability

Monica J. S. Nadler, Meredith C. Hermosura, Kazunori Inabe, Anne-Laure Perraud, Qiqin Zhu, Alexander J. Stokes, Tomohiro Kurosaki, Jean-Pierre Kinet, Reinhold Penner, Andrew M. Scharenberg & Andrea Fleig

Nature **411**, 590–595 (2001).

In this Letter, the following sentence should have appeared in the Acknowledgements section: "We gratefully acknowledge technical advice, the MerCreMer plasmid¹, and the MerCreMer transfected DT-40 cell line necessary for the creation of the Cre-Lox inducible LTRPC7 knockout cell lines from M. Reth and T. Brummer". □

1. Zhang, Y., Wienands, J., Zurn, C. & Reth, M. Induction of the antigen receptor expression on B lymphocytes results in rapid competence for signaling of SLP-65 and Syk. *EMBO J.* **17**, 7304–7310 (1998).

Coexistence of superconductivity and ferromagnetism in the d -band metal ZrZn₂

C. Pfleiderer, M. Uhlarz, S. M. Hayden, R. Vollmer, H. v. Löhneysen, N. R. Bernhoeft & G. G. Lonzarich

Nature **412**, 58–61 (2001).

In Fig. 2 the temperature axis has erroneously been multiplied by a factor of ten. These values should read 0.1 K to 0.6 K. □

The age of discovery

Lab equipment aimed at drug discovery and other high-throughput tasks.

Discovery SpeedVac

Thermo Savant www.thermo.com

A high-speed centrifugal concentrator

Designed for use in drug discovery and combinatorial chemistry, the Discovery SpeedVac is an integrated modular unit incorporating the vacuum pump and cold trap into one housing, but each component can be removed individually for maintenance. A multi-level, interchangeable rotor system provides a high throughput, with a maximum capacity of 120 shallow-well microtitre plates, 48 deep-well plates or 840 test tubes. For high-capacity sample loading, entire racks can be placed straight into the unit. Stainless steel construction with a Teflon liquid pathway allows users to run any type of solvent or solvent mixture, including aggressive acids such as TFA and organic solvents such as DMSO and DMF.

MultiMax

Mettler Toledo www.mt.com

Multiple vessel process development system

Designed for parallel organic synthesis, the MultiMax is based on a modular reactor block system equipped with four 50 ml glass reaction vessels that are temperature-controlled between -50°C and 180°C . A Windows NT-based software package allows each reaction vessel to be programmed with individual parameters for heating/cooling, stirring and dosing. Up to 16 reactors can be controlled simultaneously by one PC. MultiMax accommodates a range of analytical

methods including *in situ* reaction analysis using FTIR via ASI Applied Systems ReactIR technology. Applications include reaction screening in chemistry and optimization studies in order to speed process development in pharmaceuticals.

SlidePRO

InnoGenex www.innogenex.com

Outsourcing for localization studies

The SlidePRO service will perform customized target localization of proteins and nucleic acids for gene expression analysis of cancer, infectious disease and genetic disorders. Slide staining projects with normal, transgenic, knockout and chimaeric animal and human tissues can be carried out. Customers may select from over 500 products including detection systems, antibodies and oligonucleotide probes. High-throughput automation and reliable, reproducible staining results are promised.

Cell encapsulator

Inotech Biosystems International

www.inotechintl.com

Lock in the cells

Inotech's new cell encapsulator takes cells or other biomaterials and encases them in polymer beads or capsules. Mono-size, spherical beads can be produced in diameters ranging from $150\text{ }\mu\text{m}$ to $>1,000\text{ }\mu\text{m}$. Bead production rate varies from 200 to 5,000 beads per second, depending on the size of bead produced.

Disposable columns

Varian www.varianinc.com

No need to pack these

Designed for applications demanding high levels of purification, these new columns are built to be compatible with the new generation of automation systems. Column-packing procedures are unnecessary and the manufacturer promises improved speed, better reliability, lower solvent use and elimination of glassware. The columns are particularly suitable for use in drug development and synthesis applications involving parallel, combinatorial and other high-speed synthesis techniques.

Glutathione FlashPlate PLUS

PerkinElmer www.perkinelmer.com

Flash in the pan

This microplate is an HTS-friendly platform for the capture of glutathione-S-transferase (GST) or GST-tagged molecules.



Bound to work: FlashPlate attracts GST tags.

It is suitable for a variety of assays, including kinase and protease enzyme and protein-protein interactions. The well interiors are permanently coated with a thin layer of scintillant topped with a coating of glutathione. With no filtration steps or beads to handle, the apparatus is claimed to give consistently reproducible results with typical intra-plate CVs of 5-7%. Aspiration and wash steps can be added to improve signal-to-noise ratios. The platform can be miniaturized from 96- to 384-well density to reduce screening process and reagent costs. It is available in a 96-well format in packages of five and 20 plates, or a 384-well format in packages of two and 10 plates.

ScanTox

Harvard Apparatus

www.harvardapparatus.com

Eye-opener

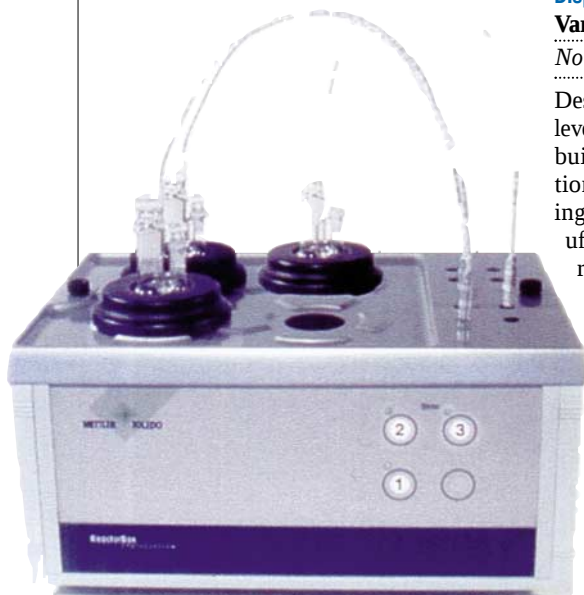
ScanTox measures the optical integrity of a cultured lens using a laser-based optical scanner. Use of the cultured lens as a measure of toxicity is a valuable tool in lens research and toxicity testing in drug discovery. Applications include pre-screening studies to test the general toxicity of chemicals and drug candidates, lens and eye research, including cataract formation and as an *in vitro* alternative to the Draize test in animals. The manufacturer points out that studies have shown direct correlations between results obtained using ScanTox and those of whole animal studies.

Multiscreen Mesh

Millipore www.millipore.com

Net a result

Each of the 96 clear, non-fluorescing, acrylic polymer wells on this plate contains an indi-



Parallel lines: MultiMax, for process development.

vidually sealed nylon mesh membrane to permit *in vivo* testing using multicellular organisms. The accompanying tray has 96 teardrop-shaped wells structured to prevent air entrapment under the membrane when the two pieces are fitted together. An optically clear tray allows for greater clarity when using microscopic analysis. The assay system is optimized for nematode *in vivo* analysis for compound and target identification and is particularly suitable for applications evaluating dosage levels of active compounds using multicellular organisms as the target for toxicity. It works by allowing the organisms to migrate through agar and mesh into the well if the target is not toxic. Dead organisms will be found on top of the agar and mesh membrane if the target is toxic.

Stacker

Radleys Discovery Technologies

www.radleys.com

A purification system that stacks up

The Stacker parallel purification system allows rapid sequential and parallel purification in a 24-well footprint. It has a compact, three-tier design that can be used to stack two levels of filtration, phase separation and/or solid-phase extraction columns, one above the other, allowing a combination of purification techniques to be used. The upper support plate is removable, making the unit suitable for single or sequential purification. For more difficult samples, a clear gas cover exerts positive pressure to increase the flow rate. The system is compatible with all standard 24-well microtitre plates as well as Greenhouse reaction blocks, Lollipop, 24 × 7 ml collection plates and other 24-position reactors/synthesizers.

MitoScan

Harvard Bioscience

www.harvardapparatus.com

Bridge the screening gap

These screening kits for late-stage discovery screening and early lead optimization are based on the mitochondrial enzymes that mediate cellular electron transport and oxidative phosphorylation. They bridge the gap between the simplicity of single-enzyme tests and the richness of response normally associated with more complex cellular and tissue assays. Processing and analysis of a 96-well microplate takes around 20 minutes. The assays are HTS-friendly, compatible with robotic sample handling and require only ml quantities. No advance preparation is needed: cell isolations and culture are not required, and there is no need for washing, extractions, filtrations or centrifugation. Suggested uses include the evaluation of mitochondrial function and as indicators of toxicity at higher levels of biological organization.

Drug discovery brochure

Pall Gelman

www.pall.com

New products and old favourites

This new brochure provides detailed information on the company's line of filtration, separation and detection products for molecular biology and drug discovery. Sample preparation and purification, PCR/labelling/probe prep, transfer and detection and high-throughput technologies are all covered. The brochure features membranes and kits for transfer and immobilization, molecular separation devices, tools for sterile filtration and venting and 96-well filter plates. Product selection criteria, performance data and ordering information are also included.

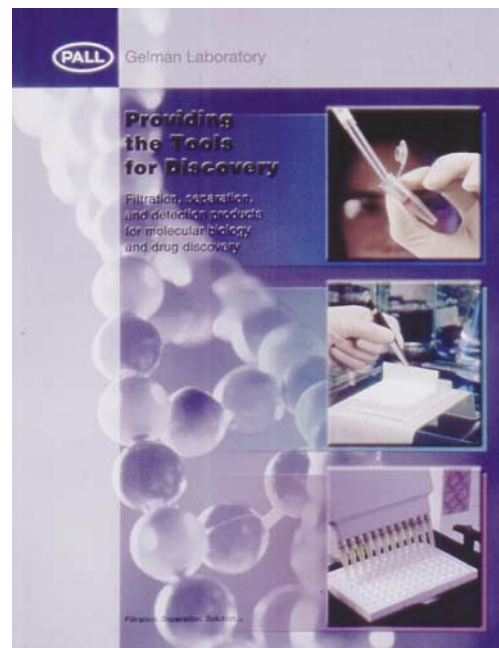
InnoGram

Innolabtech

www.innolabtech.de

Parallel synthesis system creates a stir

The InnoGram is an automated solid- and solution-phase parallel synthesis workstation for synthesis of up to 25 grams of product. The InnoGram is suitable for synthesis of intermediates or 'building blocks' in drug discovery applications. Up to eight large samples can be synthesized for further optimization synthesis or testing. Product scale ranges from 100 mg to 20 g, with up to 25 g resin for solid-phase reactions, and working reaction volumes measure 50 ml to 450 ml. InnoGram incorporates reaction vessels, temperature control from -60°C to



Be a Pall: filtration, separation and detection.

+170°C, semi-automated liquid/liquid extraction, automatic solvent and reagent addition and transfer between reaction vessels. It also features a magnetic stirring system and inert atmosphere conditions. The unit fits a standard fume hood and has a built-in PC.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

Ubiquitinylation assays?

**Ub-activating enzyme E1
Ub-conjugating enzymes
(E2s)**

**Fractions I & II
Ub conjugating kit**

www.proteasome.com

AFFINITI Research Products Ltd.
Mamhead Castle, Exeter, EX6 8HD, U.K.
T: +44(0) 1626 891010 – F +44(0) 1626 891090
– Email info@proteasome.com –

**DIVERSE
SMALL
MOLECULES
READY FOR
SCREENING**



**10K–50K
Compounds
in Plates**

- Next Day Delivery*
- Competitively Priced

*Some Restrictions Apply

Chembridge Corporation • 800.964.6143
858.451.7400 • 858.451.7401 FAX • E-mail: chem@chembridge.com

www.chembridge.com

CUSTOMISED monoclonal antibodies with GUARANTEE

BioGenes
Gesellschaft für Biopolymere mbH
Internet www.biogenes.de
Berlin *Germany
Phone +49 (0)30-65 76 23 96

BACs • YACs • PACs cDNAs

We provide clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

ResGen™
Invitrogen Corporation

**U.S. or Canada 800-533-4363
FAX 256-536-9016**

Check out our web site at
www.resgen.com

Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: classified@nature.com

Group European Manager:

Leonie Welts (4954)

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),

Andy Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Silke Opstrup (4994)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: classified@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Spain/ Portugal:

Aminda Wren

Tel + 34 91 391 3835

Fax + 34 91 319 0091

e-mail: a.wren@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: classified@nature.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

Shin-Mitsuke Building (4F),

3-6 Ichigaya Tamachi Shinjuku-Ku,

Tokyo 162

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@nature.com

Japan Manager:

Kate Cowan

Subtracting mathematicians

Hot subdisciplines, such as biostatistics, nanotechnology and systems biology, are becoming increasingly dependent on mathematics. But federal funds for maths, in the United States at least, don't look very promising. And without adequate funding, fewer people will be encouraged to go into a field that is already seeing a shrinking number of US citizens signing up (see Special Report, pages 4–5).

Not very long ago, the future for maths in the United States looked bright. Last year, Rita Colwell, director of the US National Science Foundation (NSF), suggested doubling the agency's \$120-million maths budget in two years, and quadrupling it to \$500 million over five.

Colwell made the proposal because fewer US students were entering maths programmes and fewer foreign students were filling the gaps (see *Nature* **407**, 931; 2000). One year and one president later, things have changed dramatically.

The latest suggested budget figure for the NSF's maths programme is \$140 million for the fiscal year 2002, which begins on 1 October. Although Congress is hinting that it may add some money to the NSF's budget next year, it is unlikely to be at the level Colwell has lobbied for or that the US maths community is hoping it will receive.

If the universities are unable or unwilling to back maths more heavily, mathematicians will be lured into multidisciplinary projects, for which they are very much in demand. And private-sector fields, such as finance and biostatistics, both of which are lucrative, will become more attractive to mathematicians who feel abandoned by the public sector. Perhaps after enough mathematicians leave public research, the government will decide it is worthwhile to attract replacements.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

The United States seeks the right formula to win back mathematicians

p4

WWW.NATURE.COM/ NATUREJOBS

Career centre
Information on the scientific
job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

CORBIS



Predictive power: mathematician A. N. Whitehead anticipated disappointment.

Mathematicians are in short supply in the United States. Potter Wickware examines how the federal agencies plan to boost the numbers.

The study of mathematics is apt to commence in disappointment," wrote the mathematician A. N. Whitehead in 1911, and, as if to corroborate this gloomy view, young people in the United States are turning away from the mother of the sciences. Between 1992 and 1999 the proportion of US citizens among full-time maths graduate students at the nation's top universities dropped by 26.5%, according to statistics from the National Science Foundation (NSF). Exacerbating matters, today, fewer than half of the graduate students in maths programmes across the United States are US citizens, down from 75% in the 1970s, according to American Mathematical Society data.

Strangely, the decline in interest comes just as a rich crop of theoretical and applied problems is bursting into bloom. Opportunities are abundant in cryptography, data mining, nanosystems, communications networks, materials and imaging. Physics continues to exert a strong influence on maths, with string theory revealing connections between algebraic and differential geometry. And as biology moves from a descriptive to a quantitative science, an entirely new frontier is opening up, both in physiological modelling and in new imaging systems (see "Biostatistics booms", opposite).

A decade ago, the combination of an economic recession and an influx of Russian and Chinese mathematicians made for a bleak hiring picture, but today the jobs are there, says Jim Maxwell, who tracks maths demographics at the American Mathematical Society, in Providence, Rhode Island. The current unemployment rate among new maths PhDs is 3–4%, down from over 6% in 1999. Salaries in top academic departments in 2000 ranged from \$50,000 to \$55,000 for assistant professors, \$65,000 for associates, to \$101,000 for full professors. Of the roughly 850 PhDs who enter the job market in the United States each year, about 60% take their first job in academia.

The diminishing appeal of maths to the current generation

Spooky maths

The US National Security Agency (NSA) is probably the largest employer of mathematicians in the United States. Its main mission is to make classified information impenetrable to prying eyes. A preoccupation with ciphers and keys leads to emphasis on algebra and number theory, but lots of work is also done in other areas such as signal analysis, probability theory and statistics, according to Jim Schatz, head of the agency's mathematical research group.

"We have an extremely rich set of problems here," says Schatz, who explains that restrictions on publishing in a mostly classified environment are not the constraint they might seem. There

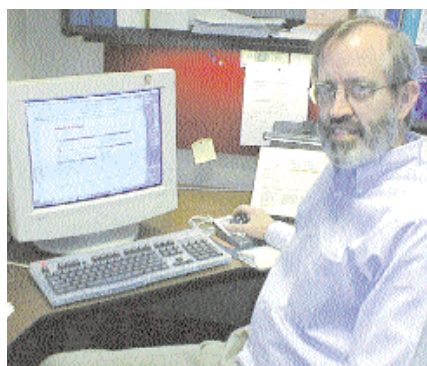
are hundreds of mathematicians in the NSA, far more than in academic settings, so readership for a typical paper is wider than in the world at large, he says.

The NSA hires about 50 mathematicians a year, about two-thirds of them PhDs. "This is a very significant percentage of US citizen maths PhDs, and we are incredibly worried about the shrinking pool," says Schatz.

To help counteract the flagging numbers of mathematicians being produced, the NSA augments the National Science Foundation's efforts with early career grants to help build the maths community. Currently about 180 grants a year, each averaging \$20,000, are being funded. **P.W.**

Booming business: Jim Maxwell believes that job prospects are good for US mathematicians.

AMS



is partly traceable to the rise of computer-science departments, which draw off those who otherwise would have majored in maths. "The dotcom world was a fantastic temptation," says Philippe Tondeur, director of the NSF's Division of Mathematical Sciences. "People with modest skills got hired without realizing how short the shelf-life of their knowledge was. With maths the skills don't go out of date."

To reverse the trend, the maths establishment is mounting an urgent if somewhat belated response. The NSF, the main patron of maths research in the United States, wants to increase its budget for maths from this year's \$120 million to between \$400 million and \$500 million by 2007. Programmes to support mathematicians at the graduate, postdoc and early career levels have been brought into place, with grants



The Mathematical Sciences Research Institute (above, left) balances pure and applied maths in class (above).

MSRI

and stipends destined to rise in number and value.

Central to the NSF's support for maths are three institutes, at Berkeley, Minneapolis and Los Angeles. Unlike the Institute for Advanced Study at Princeton, where the departments are staffed by permanent faculty, the NSF-supported maths institutes rely on 'soft' money rather than endowments and have a continually changing mathematical cast of characters.

Berkeley's Mathematical Sciences Research Institute (MSRI) dates from 1982. It has a budget of \$5 million, 70% of which comes from the NSF, with contributions from corporate partners such as Pfizer and Hewlett-Packard making up the balance. The MSRI aims for a 2:1 mix between pure and applied work, although Joe Buhler, its deputy director, is quick to point out that these terms inconveniently blur into one another. For example, number theory, once the epitome of pure maths, is now the source of many commercially important applications for computer security (see "Spooky maths", opposite).

In Minneapolis, the Institute for Mathematics and its Applications (IMA) takes a somewhat different tack, focusing on problems from industry such as networking, computational complexity and optimization. Fadil Santosa, associate director for industrial programmes, says industry is under-appreciated as a well-spring of maths jobs, naming companies such as Microsoft, Schlumberger and IBM as organizations that have maths departments. The IMA has a budget of \$3.5 million from all sources, including about \$2.2 million from the NSF.

The newest NSF-supported maths centre, the Institute for Pure and Applied Mathematics at the University of California, Los Angeles, opened its doors

this year with a five-year budget from the NSF of \$12.5 million. According to its director, Tony Chan, the institute aims to bring mathematics to new problems in the sciences. Recent programmes have been concerned with functional genomics, stochastic processes in biology and modelling of crystallization. Programmes in finance and economics focus on risk management, arbitrage, derivatives pricing in energy



Philippe Tondeur says maths skills do not date.

NSF

markets and catastrophe insurance. A third area of interest is the mathematics of image processing.

Continuing the trend of broadening the presence of maths in the scientific world, as many as four new NSF-funded institutes are in line to be created, according to Tondeur, with proposals being reviewed now. In contrast to the existing institutes, the new arrivals are likely to have more closely defined mandates, such as statistics, large data sets and imaging.

Potter Wickware is a science writer in San Francisco.

Mathematical Sciences Research Institute

♦ <http://www.msri.org>

Institute for Mathematics and its Applications

♦ <http://www.ima.umn.edu>

Institute for Pure and Applied Mathematics

♦ <http://www.ipam.ucla.edu>

MCIM



Fadil Santosa sees industry as a well-spring for maths jobs.

Biostatistics booms

Biostatistics, which is concerned with clinical-trials design and management, is booming. Demand for biostatisticians is driven by a range of issues including new drugs and therapies, testing of alternative medications, HIV/AIDS research and cancer, all fuelled by increased funding from the National Institutes of Health.

"We can't turn people out fast enough to fill the need," says Alex MacKenzie, who coordinates the biostatistics programme at the University of Washington, Seattle.

Most graduates in biostatistics are employed at universities,

or in research at pharmaceutical companies and medical research institutes. Starting salaries are \$45,000–50,000 at the masters level, and \$60,000–70,000 for PhDs, with those on the industrial side getting a moderate premium above their academic peers, says MacKenzie.

About 20 students a year graduate from Washington in biostatistics, up from six or seven a decade ago.

"Our graduates typically get flown out for interviews, are wined and dined, competed for, and usually can pick and choose among offers," says MacKenzie. **P.W.**



Tony Chan: aiming to bring maths to bear on science.

IPAM